ON THE USE OF IDEMPOTENT MATRICES IN THE TREATMENT OF LINEAR RESTRICTIONS IN REGRESSION ANALYSIS

John S. Chipman

M. M. Rao[1]

Technical Report No. 10

University of Minnesota and Carnegie Institute of Technology

University of Minnesota

Minneapolis, Minnesota

[1] Work done in part under contract Nonr 2582(00), Task NR 042-200 of the Office of Naval Research.

On the Use of Idempotent Matrices in the Treatment of Linear Restrictions in Regression Analysis[1]

by

John S. Chipman and M. M. Rao

University of Minnesota and Carnegie Institute of Technology

0. Introduction and Summary

The purpose of the present paper is to present a unified treatment of the problem of estimation and hypothesis testing in linear regression analysis. In sections 2 and 3 below, we consider the estimation of regression coefficients subject to a set of linear restrictions, according to the Markov criterion of best linear unbiasedness and the least squares criterion, respectively. In section 4 we consider the testing of a set of linear restrictions on the coefficients of a regression equation. In section 5 we take up the general linear hypothesis, that is, the testing of one set of linear restrictions subject to another set being true.

In the course of the treatment of the above problems, idempotent matrices play a natural and central role; accordingly, section 1 is devoted to the derivation of the principal theorems. Since idempotent transformations may be viewed geometrically as projections in linear spaces, section 6 is devoted to the geometric interpretation of the results. The importance of idempotent matrices in these problems was recognized by Aitken [2,3], Craig [9], Stone [27], and more recently by

---

1. Presented at the Indianapolis meetings of the Biometric Society and the American Association for the Advancement of Science, December 28, 1957; an abstract appears in Biometrics 14 (1958), 290-291. Research was conducted under Contract Nonr 2582(00), Task NR 042-200 of the Office of Naval Research.

Graybill and Marsaglia [14]; however they limited themselves to the cases in which the matrices were symmetric as well as idempotent. In section 1, unless explicitly noted to the contrary, the theorems are true for idempotent matrices that need not be symmetric.

The problems of estimation of regression coefficients subject to a set of linear restrictions, as well as that of testing a set of linear restrictions, were considered by Wilks [31] by means of a transformation of variables. The formulation presented in section 3 permits computation of the estimates directly from the ordinary (unrestricted) least-squares estimates, thus saving considerable computational work. The same is true for the problem of hypothesis testing dealt with in sections 4 and 5. The results presented here are, of course, simply a generalization of the analysis of variance and covariance; problems in this general form arise frequently in econometrics [27, 19].

1. Some properties of idempotent matrices.

An $n \times n$ matrix A is said to be idempotent if $A^2 = A$, and symmetric if $A' = A$, where $'$ denotes transposition. The trace of A, denoted tr A, is the sum of the diagonal elements a_{ii} of A. The kth-order trace of A, is the sum of all kth-order principal minors of A. The rank of A is denoted rk A. The identity matrix of order n is denoted I or I_n; the identity matrix of order $r < n$ is denoted I_r. The zero matrix is denoted 0. The symbol $\sim$ is used to denote "is distributed as". $x \sim N(\xi, \Sigma)$ means that the $n \times 1$ vector x has the multivariate normal distribution with mean vector ξ and covariance matrix Σ. $Q \sim \sigma^2\chi^2(r, \lambda)$ means that Q is distributed as σ^2 times a variable distributed as chi-square with r degrees of freedom and noncentrality parameter λ. If $\lambda = 0$ we write $Q \sim \sigma^2\chi^2(r)$.

Many of the theorems stated below are well known for the case of symmetric idempotent matrices.* The proofs of Theorems 1.3 and 1.5 are considerably simpler in that case. We do not assume symmetry, except for the case of Craig's theorem (Theorem 1.7) and its generalization. Thus we present a general form of Cochran's theorem (Theorem 1.9) in which the symmetry property is not required.

Theorem 1.1. Let A, B be $n \times n$ idempotent matrices. Then $A + B$ is idempotent if and only if $AB = BA = 0$.

Proof. The sufficiency is obvious. The necessity follows from the fact that $AB + BA = 0$, so premultiplication and postmultiplication by A gives $AB + ABA = 0 = ABA + BA$ whence $AB = BA = 0$. A straightforward generalization is contained in the

Corollary. Let $A_1, A_2, \ldots, A_k$ be $n \times n$ idempotent matrices. Then $\sum_{i=1}^{k} A_i$ is idempotent if and only if $A_iA_j = 0$ for $i \neq j$.

* Cf. Graybill and Marsaglia [14]

Theorem 1.2. Let A, B be $n \times n$ idempotent matrices. Then $A - B$ is idempotent if and only if $AB = AB = B$.

Proof. The sufficiency is obvious. To show the necessity, let $C = A - B$. Then $A = B + C$ and if C is idempotent then from Theorem 1.1, $BC = BA - B = 0$ and $CB = AB - B = 0$ so $AB = BA = B$.

Lemma 1.1. Let A be an $n \times n$ matrix of rank r. Then A may be expressed as $A = CB$ where C, B are $n \times r$ and $r \times n$ matrices, both of rank r.

Proof. Let the $n \times n$ matrix A have rank r, and let B be an $r \times n$ matrix of rank r. Then B is a basis for A, and any row a_i of A may be expressed as a linear combination $a_i = c_i B$ where c_i is a $1 \times r$ vector. Let C be the $n \times r$ matrix whose rows are such c_i; then $A = CB$ and C has rank r.

Theorem 1.3. The trace of an idempotent matrix is equal to its rank.

Proof. Let A be an $n \times n$ idempotent matrix of rank r. From Lemma 1.1, $A^2 = CBCB = CB = A$ so $BCBCBC = BCBC$, i.e., $(BC)^3 = (BC)^2$, where B, C' are $r \times n$ matrices of rank r. Since $C(BC)B = CB$ and $CB = A$ has rank r, $\operatorname{rk}(BC) \geq \operatorname{rk}(CB) = r$. But BC is $r \times r$ so $\operatorname{rk}(BC) = r$, so multiplying the equation $(BC)^3 = (BC)^2$ through by $(BC)^{-2}$ we have $BC = I_r$. Now for any $r \times n$ matrices B, C', it is easily seen that $\operatorname{tr}(BC) = \operatorname{tr}(CB)$, and so $\operatorname{tr} A = \operatorname{tr}(CB) = \operatorname{tr}(BC) = \operatorname{tr} I_r = r = \operatorname{rk} A$.

Applying this theorem to Theorems 1.1 and 1.2, and observing that for any $n \times n$ matrices A, B we have $\operatorname{tr}(A + B) = \operatorname{tr} A + \operatorname{tr} B$ and $\operatorname{tr}(A - B) = \operatorname{tr} A - \operatorname{tr} B$, we immediately have

Theorem 1.4. Let A, B be $n \times n$ idempotent matrices. Then if $AB = BA = 0$ it follows that $\operatorname{rk}(A + B) = \operatorname{rk} A + \operatorname{rk} B$, and if $AB = BA = B$ it follows that $\operatorname{rk}(A - B) = \operatorname{rk} A - \operatorname{rk} B$.

Corollary 1. Let A, B be $n \times n$ idempotent matrices of ranks r and $n - r$ respectively, such that $AB = BA = 0$; then $A = I - B$.

Corollary 2. Let A, B be $n \times n$ idempotent matrices both of rank r, such that $AB = BA = B$; then $A = B$.

We now generalize Theorem 1.3.

Theorem 1.5. Let A be an $n \times n$ idempotent matrix of rank r, and $a_k = tr_k A$ be the kth-order trace of A, i.e., the sum of the $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ kth-order principal minors of A. Then $a_k = \binom{r}{k}$ for $k \leq r$ and $a_k = 0$ otherwise.

Proof. The theorem will follow from the fact that the characteristic roots of A are all 0 or 1. For any characteristic root λ of A, let x be the corresponding characteristic vector. Then $Ax = \lambda x$ and $A^2x = \lambda^2 x$, so $\lambda x = Ax = A^2x = \lambda^2 x$, consequently $\lambda = 0$ or 1. Since this is true for all characteristic roots it follows that A has only 0 and 1 as characteristic roots.

Now if $r = n$ then $A = I_n$ so the theorem is certainly true; likewise if $r = 0$ and $A = 0$. So let $0 < r < n$. The characteristic equation of A is

$$f(\lambda) = \sum_{i=0}^{n} (-1)^i a_i \lambda^{n-i} = 0.$$

Since $r < n$, $n - r$ of the characteristic roots λ_i are zero. Now it is well known (cf. [1,p.88]) that

$$a_k = \sum_{i_1 \neq i_2 \neq \dots i_k = 1}^{r} \lambda_{i_1} \lambda_{i_2} \dots \lambda_{i_k},$$

i.e., a_k is the sum of the products of λ_i taken k at a time. Since $\lambda_i = 1$ for all non zero λ_i, and there are $\binom{r}{k}$ such terms $\lambda_{i_1} \lambda_{i_2} \dots \lambda_{i_k}$ in the above summation, it follows that $a_k = \binom{r}{k}$, proving the theorem.

We now generalize Theorems 1.1, 1.2, and 1.4.

<u>Theorem</u> 1.6. Let $A = \sum_{i=1}^{k} A_i$ be an $n \times n$ matrix of rank r, where A_i are $n \times n$ of rank r_i. Consider the following four conditions:

1) $A^2 = A$.

2) $A_i^2 = A_i$ for all i.

3) $A_iA_j = 0$ for $i \neq j$.

4) $\sum_{i=1}^{k} r_i = r$.

Then

a) 1) and 4) imply 2) and 3)

b) 2) and 3) imply 1) and 4)

c) 1) and 2) imply 3) and 4)

<u>Proof</u>. From Lemma 1.1 there exist $r_i \times n$ matrices B_i and C_i' of ranks r_i such that $A_i = C_iB_i$, for all i. We define the $n \times \Sigma r_i$ matrices

$$B' = [B_1' \quad B_2' \quad \cdots \quad B_k']$$
$$C = [C_1 \quad C_2 \quad \cdots \quad C_k].$$

Then

$$CB = \sum_{i=1}^{k} C_iB_i = \sum_{i=1}^{k} A_i = A.$$

Since A has rank r, B and C have rank $\geqq r$.

a) Let 1) and 4) hold. Since A is idempotent, $CBCB = CB$, so $BCBCBC = BCBC$, i.e., $(BC)^3 = (BC)^2$, where BC is $\Sigma r_i \times \Sigma r_i$. Since CB has rank r and $C(BC)B = CB$, BC must have rank $\geqq r$. From condition 4), $\Sigma r_i = r$ so BC is nonsingular and idempotent, hence $BC = I_r$. It follows that $B_iC_i = I_{r_i}$ and $B_iC_j = 0$ for $i \neq j$, so $A_i^2 = C_iB_iC_iB_i = C_iB_i = A_i$, and $A_iA_j = C_iB_iC_jB_j = 0$ for $i \neq j$, hence 2) and 3) hold.

b) Let 2) and 3) hold. Then $A_i^2 = C_iB_iC_iB_i = C_iB_i = A_i$, and $(B_iC_i)^3 = (B_iC_i)^2$ where B_iC_i is $r_i \times r_i$. Since $C_i(B_iC_i)B_i = C_iB_i$, $\mathrm{rk}\,(B_iC_i) \geqq \mathrm{rk}\,(C_iB_i) = r_i$, but B_iC_i is $r_i \times r_i$ so rk $(B_iC_i) = r_i$. Thus B_iC_i is nonsingular and idempotent, so $B_iC_i = I_{r_i}$. It follows that $BC = I_r$ so $A^2 = CBCB = CB = A$, therefore 1) holds. It remains to show that 4) holds. This follows from

c) Let 1) and 2) hold. Then 3) holds from the Corollary to Theorem 1.1. Since $A = \sum_{i=1}^{k} A_i$ and A, A_i are idempotent, from Theorem 1.3 we have $r = \text{rk } A = \text{tr } A = \sum_{i=1}^{k} \text{tr } A_i = \sum_{i=1}^{k} \text{rk } A_i = \sum_{i=1}^{k} r_k$ and 4) holds.

Theorem 1.7 (Craig's theorem[2]). Let the $n \times 1$ column vector x be distributed as $N(\xi, \sigma^2 I)$ and let A be an $n \times n$ symmetric matrix. Then the quadratic form $Q = x'Ax$ is distributed as $\sigma^2\chi^2(r,\lambda)$, where $\lambda = \frac{1}{2\sigma^2}\xi'A\xi$, if and only if A is idempotent of rank r.

Proof. Since A is symmetric there exists an orthogonal matrix P, with $P'P = I$, such that $PAP' = \Lambda$ where Λ is a diagonal matrix whose diagonal elements λ_i are the characteristic roots of A.

The sufficiency follows from purely algebraic considerations. If A is idempotent then $\Lambda^2 = PAP'PAP' = PAP' = \Lambda$ so Λ is also idempotent and of rank r, so contains (cf. Theorem 1.5) r 1's and $n - r$ 0's on the diagonal. Writing $z = Px$, we have $Q = x'Ax = z'PAP'z = z'\Lambda z$; without loss of generality we may write $Q = \sum_{i=1}^{r} z_i^2$. Since $x \sim N(\xi, \sigma^2 I)$, then $z \sim (\zeta, \sigma^2 I)$ where $\zeta = P\xi$. It follows from the definition of the noncentral χ^2 distribution (cf. Tang [28, p. 138]) that $Q \sim \sigma^2\chi^2(r,\lambda)$ where $\lambda = \frac{1}{2\sigma^2}\Sigma\zeta_i^2 = \frac{1}{2\sigma^2}\zeta'\Lambda\zeta = \frac{1}{2\sigma^2}\xi'A\xi$.

To show the necessity we note that $Q = z'\Lambda z = \sum_{i=1}^{n} \lambda_i z_i^2$ has a distri-

2. Cf. Craig [9].

bution with characteristic function

$$\varphi(t) = \mathcal{E}(e^{itQ}) = \mathcal{E}(e^{it\sum_{j=1}^{n}\lambda_j z_j^2}) = \prod_{j=1}^{n} (e^{it\lambda_j z_j^2})$$

$$= \prod_{j=1}^{n} (1 - 2it\sigma^2\lambda_j)^{-\frac{1}{2}} e^{-\mu_j + \mu_j(1 - 2it\sigma^2\lambda_j)^{-1}}$$

which by hypothesis is the characteristic function of $\sigma^2\chi^2(r,\lambda)$, i.e.

$$\varphi(t) = (1 - 2it\sigma^2)^{-r/2} e^{-\lambda + \lambda(1 - 2it\sigma^2)^{-1}} .$$

By the uniqueness theorem for characteristic functions (cf. Cramer [11, p.93]) it follows that $\lambda_j = 1$ for $j = 1,2,\ldots,r$ and 0 otherwise, and $\lambda = \sum_{j=1}^{n} \mu_j$, i.e., A is idempotent.

It may be remarked that since we may always write $x'Ax = x'Bx$ where B is nonsymmetric and $A = \frac{1}{2}(B + B')$ is symmetric, if A is idempotent B need not be idempotent. Conversely if B is nonsymmetric and idempotent, then $A^2 = \frac{1}{4}(B + B' + BB' + B'B) \neq A$. Since every matrix B can be written as the sum of a symmetric matrix A and a skew symmetric matrix C, where $A' = A$ and $C' = -C$, since $B^2 = B$ a small algebraic computation shows that $A^2 = A$ if and only if $C = 0$, which is impossible since $B' \neq B$. Here we can take $A = \frac{1}{2}(B + B')$, $C = \frac{1}{2}(B - B')$.

<u>Lemma</u> 1.2. Let Ω be a positive definite symmetric matrix. Then there exists a positive definite matrix T (called the square root of Ω) such that $T^2 = \Omega$, and we denote $T = \Omega^{\frac{1}{2}}$.

<u>Proof</u>. From the symmetry of Ω, there is an orthogonal transformation P such that $P'\Omega P = \Lambda$, where Λ is a diagonal matrix whose elements are real and positive (from the symmetry and positive definiteness of Ω); thus $\lambda_i^{\frac{1}{2}}$, and so $\Lambda^{\frac{1}{2}}$, exists. Denoting $T = P\Lambda^{\frac{1}{2}}P'$ we have $T^2 = P\Lambda^{\frac{1}{2}}P'P\Lambda^{\frac{1}{2}}P' = P\Lambda P' = \Omega$.

Craig's theorem is now generalized as follows (cf [14, p.684]):

<u>Theorem</u> 1.8 Let x be distributed as $N(\xi,\sigma^2\Omega)$ where Ω is a symmetric positive definite $n \times n$ matrix, and let A be a symmetric $n \times n$ matrix. Then

$Q = x'Ax$ is distributed as $\sigma^2\chi^2(r,\lambda)$, where $\lambda = \frac{1}{2\sigma^2}\,\xi'A\xi$, if and only if $A\Omega$ is idempotent of rank r.

Proof. From Lemma 1.2, $\Omega^{\frac{1}{2}}$ and $\Omega^{-\frac{1}{2}}$ exist. Let $\overset{o}{x} = \Omega^{-\frac{1}{2}}x, \mathcal{E}\overset{o}{x} = \overset{o}{\xi} = \Omega^{-\frac{1}{2}}\xi$, and let $\bar{A} = \Omega^{\frac{1}{2}}A\Omega^{\frac{1}{2}}$. Then

$$\mathcal{E}(\overset{o}{x} - \overset{o}{\xi})(\overset{o}{x}' - \overset{o}{\xi}') = \mathcal{E}\,\Omega^{-\frac{1}{2}}(x - \xi)(x' - \xi')\Omega^{-\frac{1}{2}} = \Omega^{-\frac{1}{2}}\sigma^2\Omega\Omega^{-\frac{1}{2}} = \sigma^2 I$$

so that $\overset{o}{x} \sim N(\overset{o}{\xi},\sigma^2 I)$. Now $Q = x'Ax = \overset{o}{x}'\bar{A}\overset{o}{x} = \overset{o}{x}'\Omega^{\frac{1}{2}}A\Omega^{\frac{1}{2}}\overset{o}{x} = \overset{o}{x}'\bar{A}\overset{o}{x}$, where $\bar{A}$ is symmetric since A is. From Theorem 1.7 it follows that $Q \sim \sigma^2\chi^2(r,\lambda)$, where $\lambda = \frac{1}{2\sigma^2}\,\overset{o}{\xi}'\bar{A}\overset{o}{\xi} = \frac{1}{2\sigma^2}\,\xi'A\xi$, if and only if $\bar{A}^2 = \bar{A}$, i. e., if and only if $\bar{A}^2 = \Omega^{\frac{1}{2}}A\Omega A\Omega^{\frac{1}{2}} = \Omega^{\frac{1}{2}}A\Omega^{\frac{1}{2}} = \bar{A}$. Premultiplying and postmultiplying the latter equation by $\Omega^{-\frac{1}{2}}$ and $\Omega^{\frac{1}{2}}$ respectively we obtain $A\Omega A\Omega = A\Omega$, i.e., $A\Omega$ is idempotent; thus $\bar{A}$ is idempotent if and only if $A\Omega$ is idempotent. Since Ω is positive definite(hence nonsingular), rk $(A\Omega)$ = rk A = r, completing the proof.

Corollary. Let $x \sim N(\xi,\sigma^2\Omega)$ as above, and let $y = Cx$ and $\mathcal{E}y = \mu = C\xi$, where C is an $n \times n$ idempotent matrix of rank r such that $C\Omega$ is symmetric. Then $y'\Omega^{-1}y = x'\Omega^{-1}Cx$ and $y'\Omega^{-1}y \sim \sigma^2\chi^2(r,\lambda)$, where $\lambda = \frac{1}{2\sigma^2}\,\mu'\Omega^{-1}\mu$.

Proof. $y'\Omega^{-1}y = x'C'\Omega^{-1}Cx$, where $C'\Omega^{-1}C$ is symmetric, so from Theorem 1.8, $y'\Omega^{-1}y \sim \sigma^2\chi^2(r,\lambda,)$where $\lambda = \frac{1}{2\sigma^2}\,\xi'C'\Omega^{-1}C\xi = \frac{1}{2\sigma^2}\,\mu'\Omega^{-1}\mu$, if and only if $C'\Omega^{-1}C\Omega$ is idempotent of rank r. It is certainly of rank r, since Ω is positive definite. Since $C^2 = C$ and $C\Omega = \Omega C'$, it follows that $C'\Omega^{-1}C\Omega = C'\Omega^{-1}\Omega C' = C'^2 = C'$ so $C'\Omega^{-1}C\Omega$ is idempotent. Since $C\Omega = \Omega C'$, $\Omega^{-1}C = C'\Omega^{-1}$ so $C'\Omega^{-1}C = \Omega^{-1}CC = \Omega^{-1}C$, therefore $y'\Omega^{-1}y = x'C'\Omega^{-1}Cx = x'\Omega^{-1}Cx$.

For later use we remark that the matrices A and C of Theorem 1.8 and its corollary are related by $A = \Omega^{-1}C$ and $A\Omega = C'$. In section 4 and 5 below we shall be concerned with matrices of the form C, such as the matrix $D = X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$ where X is $n \times k$ of rank k. Corresponding to the matrices A and $A\Omega$ are the matrices $\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$(which is symmetric but not idempotent) and $\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X' = D'$ (which is idempotent but not symmetric).

<u>Theorem</u> 1.9 (Cochran's theorem [3]). Let the $n \times 1$ vector x be distributed as $N(\xi, \sigma^2\Omega)$, and let $Q = x'Ax$, $Q_i = x'A_ix$, where A, A_i are $n \times n$ symmetric matrices of ranks r, r_i, and where $Q = \sum_{i=1}^{k} Q_i$. If Q is distributed as $\sigma^2\chi^2(r,\lambda)$, where $\lambda = \frac{1}{2\sigma^2}\xi' A\xi$, then a necessary and sufficient condition that the Q_i be independently distributed as $\sigma^2\chi^2(r_i,\lambda_i)$, where $\lambda_i = \frac{1}{2\sigma^2}\xi' A_i\xi$, is that (i) $r = \sum_{i=1}^{k} r_i$ or (ii) $A_i\Omega$ are idempotent for $i = 1,2,\ldots,k$.

<u>Proof</u>. Since $Q \sim \sigma^2\chi^2(r,\lambda)$, from Theorem 1.8 $A\Omega$ is idempotent. Since $Q = \sum_{i=1}^{k} Q_i$ it follows that $A = \sum_{i=1}^{k} A_i$, so $A\Omega = \sum_{i=1}^{k} A_i\Omega$. From the positive definiteness of Ω, $\text{rk}(A\Omega) = \text{rk} A = r$ and $\text{rk}(A_i\Omega) = \text{rk} A_i = r_i$. From Theorem 1.6, since condition 1) holds for $A\Omega$, conditions 4) and 2) are equivalent from a) and c) of Theorem 1.8; these are coditions (i) and (ii) stated above. Both of these imply condition 3) that $A_i\Omega A_j\Omega = 0$ for $i \neq j$, which (because of the normality of x) imply the independence of Q_i and Q_j. From Theorem 1.8, $(A_i\Omega)^2 = A_i\Omega$ if and only if $Q_i \sim \sigma^2\chi^2(r_i,\lambda_i)$, which proves the theorem.

---

3. Cf. Cochran [8].

2. Markov estimates subject to linear restrictions

In this section we summarise and generalize some well-known theorems of Gauss, Markov, and Aitken [2], which have been developed by Plackett [23], C.R. Rao[24,25], and Dwyer [13].

Lemma 2.1 (Aitken's Lemma). Let A be an $r \times n$ matrix, X a fixed $n \times k$ matrix of rank k, and Ψ a fixed $r \times k$ matrix. Then the diagonal elements of AA' reach their simultaneous minima subject to $AX = \Psi$ when

$$A = \Psi(X'X)^{-1}X'. \tag{2.1}$$

Proof. We form the Lagrangean expression

$$L = AA' - 2(AX - \Psi)\Lambda$$

where Λ is a $k \times r$ undetermined matrix. Then

$$dL = A.dA + dA.A' - 2dA.X\Lambda$$

Since $A.dA' = (dA.A')'$, the first-order condition for a minimum is

$$\text{diag}(dL) = \text{diag}(2dA[A' - X\Lambda]) = 0.$$

The second-order condition $dA.dA' > 0$ is clearly satisfied. Since dA is arbitrary it follows that $A' - X\Lambda = 0$, hence

$$A = \Lambda'X'. \tag{2.2}$$

By hypothesis, $AX = \Lambda'X'X = \Psi$, and since X is of rank k,

$$\Lambda' = \Psi(X'X)^{-1} \tag{2.3}$$

hence (2.1) follows from (2.2) and (2.3).

Definition 2.1. Let α be an $r \times 1$ vector of parameters to be estimated, and y an $n \times 1$ vector of observations. A function $a = f(y)$ is called a best linear unbiased estimator of α if

1) $a = Ay + c$

2) $\mathcal{E}a = \alpha$

3) $\mathcal{E}(a_i - \alpha_i)^2$ is a minimum for $i = 1,2,\ldots,n$.

Theorem 2.1. Let y be a random $n \times 1$ vector such that

1) $\mathcal{E}y = X\beta$

2) $\mathcal{E}(y - X\beta)(y - X\beta)' = \sigma^2\Omega$

where

a) X is a fixed and unknown $n \times k$ matrix of rank k.

b) β is a fixed and unknown $k \times 1$ vector.

c) Ω is a fixed and known $n \times n$ symmetric positive definite matrix.

d) σ^2 is a fixed and unknown scalar.

Let Ψ be a fixed and known $r \times k$ matrix. Then the best linear unbiased estimator of

$$\alpha = \Psi\beta \tag{2.4}$$

is given by

$$a = \Psi(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y. \tag{2.5}$$

Proof. Since a is a linear estimator it may be written

$$a = Ay + c. \tag{2.6}$$

Since a is an unbiased estimate of α, (2.6) and condition 1) give

$$\mathcal{E}a = AX\beta + c = \alpha = \Psi\beta \qquad \text{for all } \beta$$

hence

$$\begin{aligned} c &= 0 \\ AX &= \Psi\ . \end{aligned} \tag{2.7}$$

If $\Psi = X$ this becomes $AX = X$, hence $A(AX) = AX$ so A is an $n \times n$ idempotent matrix. Defining $\varepsilon = y - X\beta$ we have

$$a - \alpha = Ay - AX\beta = A\varepsilon,$$

so the minimum variance condition requires that the diagonal elements of

$$\mathcal{E}\,(a - \alpha)(a - \alpha)' = \mathcal{E}A\varepsilon\varepsilon'A' = A(\mathcal{E}\varepsilon\varepsilon')A' = \sigma^2A\Omega A' \tag{2.8}$$

be a minimum (using condition 2)). From condition c) it follows, using Lemma 1.2, that we may define

$$\begin{aligned} \mathring{A} &= A\Omega^{\frac{1}{2}} \\ \mathring{X} &= \Omega^{-\frac{1}{2}}X. \end{aligned} \tag{2.9}$$

Then from (2.9) and the second equation of (2.7) we have

(2.10) $\mathring{A}\mathring{X} = \Psi$

and from (2.8), (2.9) and condition d), we must minimize the diagonal terms of

(2.11) $A\Omega A' = \mathring{A}\mathring{A}'$

subject to (2.10). Applying Aitken's Lemma we obtain

$$A\Omega^{\frac{1}{2}} = \mathring{A} = \Psi(\mathring{X}'\mathring{X})^{-1}\mathring{X}' = \Psi(X'\Omega^{-\frac{1}{2}}\Omega^{-\frac{1}{2}}X)^{-1}X'\Omega^{-\frac{1}{2}}$$

so that postmultiplication by $\Omega^{-\frac{1}{2}}$ gives

(2.12) $A = \Psi(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} = \Psi B$

(defining B) whence (2.5) follows from (2.6) and the first equation of (2.7).

<u>Definition</u> 2.2. Let y be an $n \times 1$ observation vector of random variables such that $\mathcal{E}y = X\beta$ and $\mathcal{E}(y - X\beta)(y-X\beta)' = \sigma^2\Omega$, where X is a known $n \times k$ matrix of rank k, Ω a known $n \times n$ symmetric positive definite matrix of rank n, β an unknown $k \times 1$ vector of parameters, and σ^2 an unknown scalar. Let $\varepsilon = y - X\beta$. Then that value b of β for which the quadratic form $Q = \varepsilon'\Omega^{-1}\varepsilon$ is minimized for given y is called the <u>generalized least squares estimate</u> of β. The value of β which minimizes $\varepsilon'\varepsilon$ is called the (<u>simple</u>) <u>least squares</u> estimate of β.

<u>Theorem</u> 2.2. Let y, X, β, ε be as defined in Definition 2.2. Then the generalized least squares estimate of β is

(2.13) $b = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y = By.$

<u>Proof</u>. To minimize Q we set its first differential equal to zero:

$$\begin{aligned} dQ &= d(y' - \beta'X')\Omega^{-1}(y - X\beta) \\ &= - y'\Omega^{-1}X.d\beta - d\beta'.X'\Omega^{-1}y + \beta'X'\Omega^{-1}X.d\beta + d\beta'.X'\Omega^{-1}X\beta \\ &= 2d\beta'[X'\Omega^{-1}X\beta - X'\Omega^{-1}y] = 0. \end{aligned}$$

Since $d\beta'$ is arbitrary the term in brackets must vanish, giving the "normal equations"

$$X'\Omega^{-1}X\beta = X'\Omega^{-1}y.$$

Since X is of rank k the required value of β is given by (2.13). Since Ω is positive definite, from Lemma 1.2 we may write $\mathring{\varepsilon} = \Omega^{-\frac{1}{2}}\varepsilon$, so $Q = \mathring{\varepsilon}'\mathring{\varepsilon} > 0$

hence Q has no maximum. Thus (2.13) gives the value of β for which Q is a minimum.

<u>Theorem</u> 2.3. Let the conditions of Theorem 2.1 hold, and let Ψ_0 and Ψ_1 be known matrices of order $r_0 \times k$ and $r_1 \times k$ respectively, the first being of rank r_0, and let α_0 be a known $r_0 \times 1$ vector. Then the best linear unbiased estimator of

(2.14) $$\alpha_1 = \Psi_1 \beta$$

satisfying the constraint

(2.15) $$\alpha_0 = \Psi_0 \beta$$

is given by

(2.16) $$a_1 = [\Psi_1 - K\Psi_0](X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y + K\alpha_0$$

where

(2.17) $$K = \Psi_1(X'\Omega^{-1}X)^{-1}\Psi_0'[\Psi_0(X'\Omega^{-1}X)^{-1}\Psi_0']^{-1}.$$

<u>Proof</u>. From the linearity we must have

(2.18) $$a_1 = Ay + c$$

and from the unbiasedness we require the following equation

(2.19) $$\mathcal{E}a_1 = A\mathcal{E}y + c = AX\beta + c = \alpha_1 = \Psi_1\beta$$

be satisfied for all β such that $\Psi_0\beta = \alpha_0$. This may be expressed by the equation

(2.20) $$AX\beta + c - \Psi_1\beta + K(\Psi_0\beta - \alpha_0) = 0$$

where K is an $r_1 \times r_0$ undetermined matrix of full rank. Collecting terms, (2.20) becomes

(2.21) $$[AX - \Psi_1 + K\Psi_0]\beta = -c + K\alpha_0$$

identically in β, whence

(2.22) $$\begin{aligned} c &= K\alpha_0 \\ AX &= \Psi_1 - K\Psi_0. \end{aligned}$$

If $\Psi_1 = X$ this becomes $Ax = X - K\Psi_0$; since (2.22) must hold for any Ψ_0 including the empty matrix $\Psi_0 = \emptyset$, we have $AAX = AX$ so A is an $n \times n$ idempotent matrix. We now form the expression

(2.23) $L = A\Omega A' - 2[AX - \Psi_1 + K\Psi_0]\Lambda - 2[c - K\alpha_0]\ \lambda'$

where Λ and λ' are $k \times r_1$ and $1 \times r_1$ matrices of Lagrangean multipliers. The diagonal elements of the first differential of L are given by

(2.24) $\operatorname{diag}(dL) = \operatorname{diag}\{2dA[\Omega A' - X\Lambda] - 2dc.\lambda' - 2dK[\Psi_0\lambda - K\alpha_0\Lambda']\}.$

Since dA, dc, and dK are arbitrary, this leads to

$$\Omega A' - X\Lambda = 0$$

(2.25) $$\lambda' = 0$$

$$\Psi_0 - K\alpha_0\lambda' = \Psi_0\ \Lambda = 0.$$

Since Ω is positive definite the second-order condition $\operatorname{diag}(d^2L) = 2dA\Omega dA' = 2dA\Omega^{\frac{1}{2}}\Omega^{\frac{1}{2}}dA' > 0$ is satisfied. From the first equation of (2.25) and the nonsingularity of Ω we have

(2.26) $A = \Lambda'X'\Omega^{-1}.$

From (2.26) and the second equation of (2.22) we have

(2.27) $\Lambda'X'\Omega^{-1}X = \Psi_1 - K\Psi_0$

and since X is of rank k this gives

(2.28) $\Lambda' = [\Psi_1 - K\Psi_0](X'\Omega^{-1}X)^{-1}.$

Thus we have

(2.29) $A = [\Psi_1 - K\Psi_0](X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$

which, together with the first equation of (2.22), establishes (2.16). It remains to evaluate K.

Now postmultiplying (2.27) by $(X'\Omega^{-1}X)^{-1}\Psi_0'$ and using the third equation of (2.25) we obtain

(2.30) $K\Psi_0(X'\Omega^{-1}X)^{-1}\Psi_0' = \Psi_1(X'\Omega^{-1}X)^{-1}\Psi_0'$

and since Ψ_0 is of rank r_0, (2.17) follows.

We may finally note that, defining $G_0 = (X'\Omega^{-1}X)^{-1}\Psi_0'[\Psi_0(X'\Omega^{-1}X)^{-1}\Psi_0']^{-1}$, we have $K = \Psi_1G_0$ and (2.28) may be written

(2.31) $\Lambda' = \Psi_1[I - G_0\Psi_0](X'\Omega^{-1}X)^{-1},$

where $G_0\Psi_0$ is idempotent of rank r_0 and $I - G_0\Psi_0$ is idempotent of rank $n - r_0$. If $\Psi_1 = X$ then the matrix $A = X[I - G_0\Psi_0](X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$ is idempotent of rank $n - r_0$.

3. Least squares estimates subject to linear restrictions.

The main result of this section is given by the following

Theorem 3.1 Let y be an $n \times 1$ column vector of random variables such that

$$(3.1)\qquad y = X\beta + \varepsilon$$

where X is an $n \times k$ matrix (of rank k) of known fixed variates, β is a $k \times 1$ vector of unknown parameters, and ε is an $n \times 1$ vector of random variables with $\mathcal{E}\varepsilon = 0$ and $\mathcal{E}\varepsilon\varepsilon' = \sigma^2\Omega$ where Ω is a known positive definite symmetric matrix and σ^2 an unknown positive scalar. Furthermore let

$$(3.2)\qquad \Psi\beta = \alpha$$

where Ψ is a known $r \times k$ matrix of rank k, and α is a known $r \times 1$ vector. Then the generalized least squares estimate of β subject to (3.2) is given by

$$(3.3)\qquad \hat{\beta} = b - (X'\Omega^{-1}X)^{-1}\Psi'[\Psi(X'\Omega^{-1}X)^{-1}\Psi']^{-1}(\Psi b - \alpha)$$

where b is the unrestricted generalized least squares estimate of β given by

$$(3.4)\qquad b = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$$

Proof. Since Ψ is of rank r, it is possible to find an $l \times k$ matrix $\tilde{F}$, where $l = k - r$, such that the matrix

$$(3.5)\qquad T' = [\tilde{F}' \quad \Psi']$$

is nonsingular.[4] We adopt the notation

$$(3.6)\qquad T^{-1} = [\Phi \quad \tilde{G}]$$

for a partition of T^{-1} into its first l and last r columns.[5]

---

4. The matrix $\tilde{F}$ is, of course, not unique. A unique representation may be obtained from the following lexicographic criterion. Let $\tilde{F}_\kappa = [\delta^{k1}, \delta^{k2}, \ldots \delta^{kl}]$, where δ^{ki} is a k th - order column vector with 1 in the k_i th place and zeros elsewhere. Let $\kappa = [k_1, k_2, \ldots k_l]$ be a row vector, lexicographically ordered by the definition $\kappa >^L \kappa^*$ if and only if, for all j such that $k_j < k_j^*$ there exists an $i < j$ such that $k_i > k_i^*$. Then we choose $\tilde{F}_\kappa$ such that κ is a maximum.

5. Compare Koopmans, Rubin and Leipnik [19, p. 160]. Ψ and Φ correspond to what they call the restriction matrix and the basic matrix respectively.

Then

$$(3.7)\qquad TT^{-1} = \begin{bmatrix} \tilde{F} \\ \Psi \end{bmatrix} [\Phi \quad \tilde{G}] = \begin{bmatrix} \tilde{F}\Phi & \tilde{F}\tilde{G} \\ \Psi\Phi & \Psi\tilde{G} \end{bmatrix} = \begin{bmatrix} I_l & 0 \\ 0 & I_r \end{bmatrix}$$

and

$$(3.8)\qquad T^{-1}T = [\Phi \quad \tilde{G}] \begin{bmatrix} \tilde{F} \\ \Psi \end{bmatrix} = \Phi\tilde{F} + \tilde{G}\Psi = \tilde{L} + \tilde{R} = I_k$$

From (3.7) it is easily seen that the matrices $\tilde{L} = \Phi\tilde{F}$ and $\tilde{R} = \tilde{G}\Psi$ of (3.8) are idempotent of ranks l and r respectively, and orthogonal complements; i.e., $\tilde{L}^2 = \tilde{L}$, $\tilde{R}^2 = \tilde{R}$, and $\tilde{L}\tilde{R} = \tilde{R}\tilde{L} = 0$.

Now we define the transformations[6]

$$(3.9)\qquad \tilde{\beta} = T\beta = \begin{bmatrix} \tilde{F}\beta \\ \Psi\beta \end{bmatrix} = \begin{bmatrix} \beta^* \\ \alpha \end{bmatrix}$$

$$(3.10)\qquad \tilde{X} = XT^{-1} = [X\Phi \quad X\tilde{G}].$$

Then β may be written

$$(3.11)\qquad \beta = T^{-1}\tilde{\beta} = \Phi\beta^* + \tilde{G}\alpha = \tilde{L}\beta + \tilde{R}\beta$$

and (3.1) becomes

$$(3.12)\qquad y = \tilde{X}\tilde{\beta} + \varepsilon = X\Phi\beta^* + X\tilde{G}\alpha + \varepsilon.$$

Defining

$$(3.13)\qquad y^* = y - X\tilde{G}\alpha$$

$$(3.14)\qquad X^* = X\Phi$$

we may replace (3.12) by

$$(3.15)\qquad y^* = X^*\beta^* + \varepsilon.$$

Since (3.15) incorporates the restrictions (3.2), we seek the generalized least squares estimate of β^* in (3.15) which is

$$(3.16)\qquad b^* = (X^{*\prime}\Omega^{-1}X^*)^{-1}X^{*\prime}\Omega^{-1}y^* = F(b - \tilde{G}\alpha)$$

where use is made of (3.13), (3.14), and (3.4), and where

$$(3.17)\qquad F = (X^{*\prime}\Omega^{-1}X^*)^{-1}X^{*\prime}\Omega^{-1}X = (\Phi'X'\Omega^{-1}X\Phi)^{-1}\Phi'X'\Omega^{-1}X.$$

---

6. Compare Wilks [31, pp. 166-175].

This matrix F is known as the "alias matrix";[7] it also plays a central part in linear aggregation analysis in econometrics (cf. Theil [29,30]).

Now in order to obtain the desired estimator of β, we substitute in (3.11) the estimate b* given by (3.16) for β^*, obtaining

(3.18) $$\hat{\beta} = \Phi b^* + \tilde{G}\alpha = L(b - \tilde{G}\alpha) + \tilde{G}\alpha = Lb + R\tilde{G}\alpha = b - R(b - \tilde{G}\alpha)$$

where we define

(3.19) $$L = \Phi F = I - R.$$

(3.18) expresses $\hat{\beta}$ in terms of the undetermined matrices Φ and $\tilde{G}$, whereas it is more advantageous to have an expression for $\hat{\beta}$ in terms of the known matrix Ψ. Now it is easily verified from (3.17) that L is a $k \times k$ idempotent matrix of rank l, hence by Theorems 1.2 and 1.4, R is idempotent of rank r, and furthermore $LR = RL = 0$. Making use of (3.7), R must then be of the form

(3.20) $$R = (X'\Omega^{-1}X)^{-1}\Psi' S^{-1}\Psi$$

where S is some $r \times r$ nonsingular matrix. Since R is idempotent,

(3.21) $$RR = (X'\Omega^{-1}X)^{-1}\Psi' S^{-1}\Psi(X'\Omega^{-1}X)^{-1}\Psi' S^{-1}\Psi = (X'\Omega^{-1}X)^{-1}\Psi' S^{-1}\Psi = R$$

which can be true only if

(3.22) $$S = \Psi(X'\Omega^{-1}X)^{-1}\Psi'$$

hence

(3.23) $$R = (X'\Omega^{-1}X)^{-1}\Psi'[\Psi(X'\Omega^{-1}X)^{-1}\Psi']^{-1}\Psi = G\Psi$$

where

(3.24) $$G = (X'\Omega^{-1}X)^{-1}\Psi'[\Psi(X'\Omega^{-1}X)^{-1}\Psi']^{-1}.$$

Now substitution of (3.23) in (3.18) gives, after making use of (3.7),

(3.25) $$\hat{\beta} = b - G(\Psi b - \alpha) = (I - G\Psi)b + G\alpha = Lb + G\alpha$$

which establishes (3.3), hence the theorem is proved.

---

7. Cf. Box and Wilson [6, p. 7] and Box and Hunter [5, p. 198].

4. Test of linear restrictions.

The principal result of this section is contained in the following theorem.

Theorem 4.1. Let y be distributed as $N(X\beta, \sigma^2\Omega)$, where X is a known $n \times k$ matrix of rank $k < n$, Ω a known $n \times n$ positive definite symmetric matrix, β an unknown $k \times 1$ vector, and σ^2 an unknown positive scalar. Let Ψ be a known $r \times k$ matrix of rank r, and α a known $r \times 1$ vector. Then an unbiased critical region for testing the hypothesis $\Psi\beta = \alpha$ at significance level θ, against the alternative hypothesis $\Psi\beta \neq \alpha$, is

$$(4.1) \quad F_{r,n-k} = \frac{n-k}{r} \cdot \frac{(b'\Psi' - \alpha')[\Psi(X'\Omega^{-1}X)^{-1}\Psi']^{-1}(\Psi b - \alpha)}{y'\Omega^{-1}y - y'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y} > F_{r,n-k}(\theta)$$

where $b = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ is the generalized least squares estimate of β and $F_{r,n-k}(\theta)$ is the upper θ percentage point of the F distribution with r and $n - k$ degrees of freedom.

Proof. We define the residual $n \times 1$ vectors

$$(4.2) \quad e = y - Xb$$

$$(4.3) \quad \hat{\varepsilon} = y - X\hat{\beta} = y - Xb + XG(\Psi b - \alpha)$$

where b and $\hat{\beta}$ are defined by (3.4) and (3.25) respectively, and the $n \times 1$ vectors

$$(4.4 \quad h = \hat{\varepsilon} - e = XG(\Psi b - \alpha) = X(b - \hat{\beta})$$

$$(4.5) \quad z = y - XG\alpha$$

with means respectively

$$(4.6) \quad \eta = \mathcal{E}h = XG(\Psi\beta - \alpha)$$

$$(4.7) \quad \zeta = \mathcal{E}z = X(\beta - G\alpha)$$

Note that if we take $G = \tilde{G}$, then $z = y^*$.

Defining the $n \times n$ matrices

$$(4.8) \quad D = X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} = I - E$$

$$(4.9) \quad H = X(X'\Omega^{-1}X)^{-1}\Psi'[\Psi(X'\Omega^{-1}X)^{-1}\Psi']^{-1}\Psi(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$$

it is seen that

$$(4.10) \quad Ez = Ey = y - Xb = e$$

(4.11) $\qquad Hz = XG(\Psi b - \alpha) = h$;

and, since $EX = 0$ and $HX = XG\Psi$,

(4.12) $\qquad E\zeta = EX(\beta - G\alpha) = 0$

(4.13) $\qquad H\zeta = HX(\beta - G\alpha) = \eta$.

It is easily verified that D and H are idempotent of ranks k and r respectively, hence E is idempotent of rank $n - k$ by Theorems 1.2 and 1.4. Furthermore $DH = HD = H$ so that $EH = HE = 0$, consequently the $n \times n$ matrix

(4.14) $\qquad \hat{E} = E + H = I - D + H = I - \hat{D}$

is idempotent of rank $n - k + r$ by Theorems 1.1 and 1.4, and

(4.15) $\qquad \hat{E}z = (E + H)z = e + h = \hat{\varepsilon}$

from (4.10), (4.11), and (4.4).

Now z is distributed as $N(\zeta, \sigma^2\Omega)$, hence from (4.10), using Theorem 1.8, $e'\Omega^{-1}e$ is distributed as $\sigma^2\chi^2(n - k)$ since E is idempotent of rank $n - k$ and $\mathcal{E}\varepsilon = 0$. Likewise, from (4.11) $h'\Omega^{-1}h$ is distrbuted as $\sigma^2\chi^2(r,\lambda)$, since H is idempotent of rank r and

(4.16) $\qquad \lambda = \dfrac{1}{2\sigma^2}\,\eta'\Omega^{-1}\eta \geqq 0.$

Furthermore from (4.5) and (4.7) we have $z - \zeta = \varepsilon$, and since $Hz = h$ and $H\zeta = \eta$ from (4.11) and (4.13), $H\varepsilon = h - \eta$. Similarly $E\varepsilon = E(z - \zeta) = Ez = e$. Since $HE = EH = 0$, the covariance between e and h (which have singular normal distributions) is

(4.17) $\quad \mathcal{E}(h - \eta)e' = \mathcal{E}H\varepsilon\varepsilon'E' = H(\mathcal{E}\varepsilon\varepsilon')E' = H\sigma^2\Omega E' = \sigma^2 HE\Omega = 0$

and similarly $\mathcal{E}e(h - \eta)' = 0$; therefore $e'\Omega^{-1}e$ and $h'\Omega^{-1}h$ are independently distributed by central and noncentral chi-square laws respectively. Hence the variance ratio

(4.18) $\qquad F_{r,n-k} = \dfrac{n - k}{r} \cdot \dfrac{h'\Omega^{-1}h}{e'\Omega^{-1}e} = \dfrac{n - k}{r} \cdot \dfrac{z'\Omega^{-1}Hz}{z'\Omega^{-1}Ez}$

has the noncentral F distribution with r and $n - k$ degrees of freedom and noncentrality λ as given by (4.16).

To establish (4.1) we observe from (4.4) that

(4.19) $h'\Omega^{-1}h = (b'\Psi' - \alpha')G'X'\Omega^{-1}XG(\Psi b - \alpha) = (b'\Psi' - \alpha')[\Psi(X'\Omega^{-1}X)^{-1}\Psi']^{-1}$

$(\Psi b - \alpha)$ and from (4.2) that

(4.20) $e'\Omega^{-1}e = y'[\Omega^{-1} - \Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y] = y'\Omega^{-1}y - y'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y.$

It remains to establish the unbiasedness of the test.

The unbiasedness will be established following closely an argument of R.D.Narain [22]. From (4.18) the critical region W is defined by

(4.21) $W = \{ F_{r,n-k} = \frac{n-k}{r} \frac{h'\Omega^{-1}h}{e'\Omega^{-1}e} \mid F_{r,n-k} > F_{r,n-k}(\theta) \}$

where $F_{r,n-k}(\theta)$ is a constant. The complement of this region, say W', is given by

(4.22) $W' = \{ F_{r,n-k} \mid \frac{h'\Omega^{-1}h}{e'\Omega^{-1}e} \leq C \}$

where

(4.23) $C = \frac{r}{n-k} F_{r,n-k}(\theta)$

or equivalently,

(4.24) $W' = \{ F_{r,n-k} \mid \chi^2(r,\lambda) \leq C\chi^2(n-k) \}$

where $\chi^2(r,\lambda) = \frac{1}{\sigma^2} h'\Omega^{-1}h$ and $\chi^2(n-k) = \frac{1}{\sigma^2} e'\Omega^{-1}e$. If $\psi(\lambda) = \Pr\{ W' \}$ then $\psi(\lambda)$ is the probability of type II error (i.e., $1 - \psi(\lambda)$ is the power), and we wish to show that $\psi(\lambda) \leq \psi(0)$ which is what is meant by unbiasedness of the test.

For convenience let χ'^2_r stand for the noncentral χ^2 with r degrees of freedom. Since χ'^2_r and χ^2_{n-k} are independent, $\psi(\lambda)$ can be found by first obtaining the conditional probability of W' given χ^2, and then integrating out χ^2(i.e., $\psi = \mathcal{E}[\mathcal{E}(\iota_{W'} \mid \chi^2)]$ where $\iota_{W'}$ is the indicator random variable defined by $\iota_{W'} = 1$ if $\chi'^2 \epsilon W'$ and $\iota_W = 0$ if $\chi'^2 \not\epsilon W'$). Hence

(4.24) $$\psi(\lambda) = \int_0^\infty \left[\int_0^{C\chi^2_{n-k}} f(\chi'^2_r)\, d\chi'^2 \right] f(\chi^2_{n-k})\, d\chi^2$$

where $f(\chi'^2)$ and $f(\chi^2)$ are the respective densities of χ'^2 and χ^2. Substituting the expression for $f(\chi'^2)$ in (4.24) we obtain

$$(4.25) \qquad \psi(\lambda) = \int_0^\infty \left[\int_0^{C^2_{n-k}} \frac{e^{-\chi'^2/2}\, e^{-\lambda/2}}{2^{r/2}} \sum_{j=0}^\infty \frac{(\chi'^2)^{2/2+j-1}\, \lambda^j}{2^{2j} j!\, (r/2+j)}\, d\chi'^2 \right] f(\chi^2)\, d\chi^2$$

$$= \frac{e^{-\lambda/2}}{2^{r/2}} \sum_{j=0}^\infty \frac{\lambda^j}{2^{2j} j!\, \Gamma(r/2+j)} \int_0^\infty \left[\int_0^{C_{n-k}} e^{-\chi'^2/2} (\chi'^2)^{r/2+j-1}\, d\chi'^2 \right] f(\chi^2)\, d\chi^2,$$

the interchange of integration and summation being trivially justified, since the series is one of nonnegative terms, and is convergent. To show that $\psi(\lambda)$ is decreasing as λ increases, we find that

$$(4.26) \qquad \frac{d\psi(\lambda)}{d\lambda} < 0 \quad \text{for all } \lambda > 0$$

(differentiation term by term is again a trivial matter to justify). From (4.26) it follows that $\psi(\lambda)$ is decreasing and that

$$(4.27) \qquad \psi(\lambda) \leqq \psi(0) \quad \text{for all } \lambda \geqq 0$$

which completes the proof of the theorem.

An argument similar to Narain's has also been used by K.V.Ramachandran [26]. In Theorem 5.1 below, the unbiasedness condition will also be stated, but the proof, being completely similar to the above, will not be repeated.

5. Test of one set of linear restrictions subject to another set being true.

5.1 In this section the results of the previous section are generalized by the following theorem:

Theorem 5.1 Let the $n \times 1$ vector y be distributed as $N(X\beta, \sigma^2\Omega)$, where X is a fixed and known $n \times k$ matrix of rank k, Ω a fixed and known positive definite symmetric matrix, β an unknown $k \times 1$ vector of parameters, and σ^2 an unknown positive scalar. Let Ψ_0, Ψ_1 be known $r_0 \times k$ and $r_1 \times k$ matrices of ranks r_0 and r_1 respectively, such that $\Psi' = [\Psi_1' \;\; \Psi_0']$ has rank $r_1 + r_0 = r \leq k$; and let α_0 and α_1 be known $1 \times r_0$ and $1 \times r_1$ vectors, and $\alpha' = [\alpha_1' \;\; \alpha_0']$. Then an unbiased critical region of size θ for testing the hypothesis $\Psi_1\beta = \alpha_1$ against the alternative hypothesis $\Psi_1\beta \neq \alpha_1$, subject to the hypothesis $\Psi_0\beta = \alpha_0$ being true, is given by

$$(5.1) \qquad F_{r_1, n-k+r_0} = \frac{n-k+r_0}{r_1} \cdot \frac{c'S^{-1}c - c_0'S_{00}^{-1}c_0}{e'\Omega^{-1}e + c_0'S_{00}^{-1}c_0} > F_{r_1, n-k+r_0}(\theta)$$

where $F_{r_1, n-k+r_0}(\theta)$ is the upper θ percentage point of the F distribution with r_1 and $n - k + r_0$ degrees of freedom, and where

$$(5.2) \qquad S = \Psi(X'\Omega^{-1}X)^{-1}\Psi' = \begin{bmatrix} \Psi_1(X'\Omega^{-1}X)^{-1}\Psi_1' & \Psi_1(X'\Omega^{-1}X)^{-1}\Psi_0' \\ \Psi_0(X'\Omega^{-1}X)^{-1}\Psi_1' & \Psi_0(X'\Omega^{-1}X)^{-1}\Psi_0' \end{bmatrix} = \begin{bmatrix} S_{11} & S_{10} \\ S_{01} & S_{00} \end{bmatrix}$$

and

$$(5.3) \qquad c = \Psi b - \alpha = \begin{bmatrix} \Psi_1 b - \alpha_1 \\ \Psi_0 b - \alpha_0 \end{bmatrix} = \begin{bmatrix} c_1 \\ c_0 \end{bmatrix}$$

$$e = y - Xb$$

where $b = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$ is the generalized least square estimate of β.

Proof. The generalized least squares estimates of β subject to $\Psi\beta = \alpha$ and $\Psi_0\beta = \alpha_0$ are given respectively by

$$(5.4) \qquad \begin{aligned} \hat{\beta} &= b - G(\Psi b - \alpha) = b - (X'\Omega^{-1}X)^{-1}\Psi'[\Psi(X'\Omega^{-1}X)^{-1}\Psi']^{-1}(\Psi b - \alpha) \\ \beta^0 &= b - G_0(\Psi_0 b - \alpha_0) = b - (X'\Omega^{-1}X)^{-1}\Psi_0'[\Psi_0(X'\Omega^{-1}X)^{-1}\Psi_0']^{-1}(\Psi_0 b - \alpha_0) \end{aligned}$$

from (3.24). We define the matrices

$$(5.5) \quad Z = \begin{bmatrix} 0 & 0 \\ 0 & S_{00}^{-1} \end{bmatrix}; \quad W = S^{-1} - Z$$

where S and S_{00} are defined by (5.2). We note for later use that

$$(5.6) \quad ZSZ = Z; \quad WSW = W.$$

We now define

$$\begin{aligned} G &= (X'\Omega^{-1}X)^{-1}\Psi'S^{-1} \\ (5.7) \quad G^0 &= (X'\Omega^{-1}X)^{-1}\Psi'Z \\ G\# &= (X'\Omega^{-1}X)^{-1}\Psi'W = G - G^0 \end{aligned}$$

Then since $\Psi' = [\Psi_1' \ \ \Psi_0']$ and $\alpha' = [\alpha_1' \ \ \alpha_0']$, we may write $\beta^0 = b - G^0(\Psi b - \alpha)$, hence from (5.4) we have

$$\begin{aligned} b - \hat{\beta} &= G(\Psi b - \alpha) = Rb - G\alpha \\ (5.8) \quad b - \beta^0 &= G^0(\Psi b - \alpha) = R_0 b - G^0\alpha \\ \beta^0 - \hat{\beta} &= G\#(\Psi b - \alpha) = R\# b - G\#\alpha \end{aligned}$$

The above equations define $R = G\Psi$; $R_0 = G^0\Psi = G_0\Psi_0$, and $R\# = G\#\Psi = G\Psi - G_0\Psi_0$.

We now define the residual vectors

$$\begin{aligned} e &= y - Xb \\ (5.9) \quad \varepsilon^0 &= y - X\beta^0 \\ \hat{\varepsilon} &= y - X\hat{\beta} \end{aligned}$$

and the vectors

$$\begin{aligned} h &= \hat{\varepsilon} - e = X(b - \hat{\beta}) = XG(\Psi b - \alpha) \\ (5.10) \quad h^0 &= \varepsilon^0 - e = X(b - \beta^0) = XG^0(\Psi b - \alpha) \\ h\# &= \hat{\varepsilon} - \varepsilon^0 = X(\beta^0 - \hat{\beta}) = XG\#(\Psi b - \alpha) \end{aligned}$$

Then $h\# = h - h^0$ and $\varepsilon^0 = e + h^0$. We shall proceed to show that the quadratic forms $h\#'\Omega^{-1}h\#$ and $\varepsilon^{0}{}'\Omega^{-1}\,{}^0$ are independently distributed by non-central and central chi-square laws respectively.

As in the previous section we define

$$D = X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} = I - E$$

$$(5.11) \quad H = X(X'\Omega^{-1}X)^{-1}\Psi'S^{-1}\Psi(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$$

$$H_0 = X(X'\Omega^{-1}X)^{-1}\Psi_0'S_{00}^{-1}\Psi_0(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$$

and we further define

$$(5.12) \quad D_0 = D - H_0 = I - E_0$$

$$E\# = H - H_0 = \hat{E} - E_0 \ .$$

Note that, using the definitions of R, R_0, $R\#$ in (5.8),

$$H = X(X'\Omega^{-1}X)^{-1}\Psi'S^{-1}\Psi(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} = XRB$$

$$(5.13) \quad H_0 = X(X'\Omega^{-1}X)^{-1}\Psi'Z\Psi(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} = XR_0B$$

$$H\# = X(X'\Omega^{-1}X)^{-1}\Psi'W\Psi(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} = XR\#B$$

where

$$(5.14) \quad B = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}.$$

From (5.11) we note that

$$(5.15) \quad DH = H = HD \ ; \ EH = 0 = HE$$

$$DH_0 = H_0 = H_0D \ ; \ EH_0 = 0 = H_0E \quad .$$

Consequently, since D, H, and H_0 are idempotent, simple computation shows that D_0 is idempotent of rank $k - r_0$, from Theorems 1.2 and 1.4.

Similarly, simple computation, making use of the definitions of S and Z in (5.2) and (5.5), shows that

$$G\Psi G = G = G\Psi G$$

$$(5.16) \quad G\Psi G^0 = G^0 = G^0\Psi G$$

$$G\Psi G\# = G\# = G\#\Psi G$$

from which it follows immediately that

$$RR = R = RR$$

$$(5.17) \quad RR_0 = R_0 = R_0R$$

$$RR\# = R\# = R\#R,$$

and similarly (using the last equalities in (5.13) and the property $BX = I$)

$$HH = H = HH$$

(5.18) $$HH_0 = H_0 = H_0H$$

$$HH\# = H\# = H\#H$$

Consequently $H\#$ is idempotent, of rank $r - r_0 = r_1$, from Theorems 1.2 and 1.4.

As in (4.5) and (4.7) we define

(5.19) $$z = y - XG\alpha$$

$$\mathcal{E}z = \zeta = X\beta - XG\alpha$$

so that

(5.20) $$z - \zeta = y - X\beta = \varepsilon \ .$$

Now from (5.19), (5.10), and (5.13) it is easily seen using (5.16) that

$$Hz = h$$

(5.21) $$H_0z = h^0$$

$$H\#z = (H - H_0)z = h - h^0 = h\#$$

$$E_0z = (E + H_0)z = e + h^0 = \varepsilon^0 \ .$$

We also have

$$\mathcal{E}h = H\zeta = \eta$$

(5.22) $$\mathcal{E}h^0 = H_0\zeta = \eta^0$$

$$\mathcal{E}h\# = H\#\zeta = \eta\#$$

$$\mathcal{E}\varepsilon^0 = E_0\zeta = 0 \ ,$$

the first three equations defining η, η^0, $\eta\#$, and the last equation following from the fact that $EX = 0$ and $H_0\zeta = XG^0(\Psi\beta - \alpha) = XG_0(\Psi_0\beta - \alpha_0) = 0$, where use is made of (5.16) and the maintained hypothesis $\Psi_0\beta = \alpha_0$.

From (5.20), (5.21), (5.22) we have

(5.23) $$H\#\varepsilon = H\#(z - \zeta) = h\# - \mathcal{E}h\# = h\# - \eta\#.$$

$$E_0\varepsilon = E_0(z - \zeta) = \varepsilon^0 - \mathcal{E}\varepsilon^0 = \varepsilon^0 \ .$$

We may also note that

(5.24) $$H\#E_0 = (H - H_0)(E + H_0) = 0,$$

making use of (5.15) and (5.18), and similarly $E_0H\# = 0$. Thus $H\#$ and E_0 are orthogonal idempotent matrices, of ranks r_1 and $n - k + r_0$ respectively, so

$E_0 + H\# = E + H = \hat{E}$ is idempotent of rank $n - k + r_0 + r_1 = n - k + r$ by Theorems 1.1 and 1.4.

Now $\varepsilon = y - X\beta$ is distributed as $N(0, \sigma^2\Omega)$, since $y \sim N(X\beta, \sigma^2\Omega)$. So, making use of (5.23) and (5.24), and the fact that $E_0' = \Omega^{-1}E_0\Omega$, we have for the covariance between $h\#$ and ε^0:

$$(5.25) \quad \mathcal{E}(h\# - \eta\#)\varepsilon^{0\prime} = \mathcal{E}H\#\,\varepsilon\varepsilon' E_0' = H\#(\mathcal{E}\varepsilon\varepsilon')E_0' = H\#\sigma^2\Omega E_0' = \sigma^2 H\# E_0\Omega = 0 .$$

Similarly, $\mathcal{E}\varepsilon^0(h\#' - \eta\#') = 0$. Thus $h\#$ and ε^0, which have singular normal distributions, are uncorrelated and therefore independently distributed. It follows that $h\#'\Omega^{-1}h\#$ and $\varepsilon^{0\prime}\Omega^{-1}\varepsilon^0$ are independently distributed; we shall now show that they are also distributed by χ^2 laws.

Since z is a translate of y by (5.19), it is distributed as $N(\zeta, \sigma^2\Omega)$, so from Theorem 1.8, $\varepsilon^{0\prime}\Omega^{-1}\varepsilon^0 = z'E_0'\Omega^{-1}E_0 z = z'\Omega^{-1}E_0 z$ is distributed as $\sigma^2\chi^2(n - k + r_0)$, since the noncentrality parameter is proportional to $\zeta'\Omega^{-1}E_0\zeta = 0$ because $E_0\zeta = 0$ from (5.22). Likewise $h\#'\Omega^{-1}h\# = z'\Omega^{-1}H\#z$ is distributed as $\sigma^2\chi^2(r_1, \lambda)$ where

$$(5.26) \quad \lambda = \frac{1}{2\sigma^2}\,\zeta'\Omega^{-1}H\#\zeta = \frac{1}{2\sigma^2}\,\eta\#'\Omega^{-1}\eta\# = \frac{1}{2\sigma^2}\mathring{\eta}\#'\mathring{\eta}\# \geq 0$$

and where $\mathring{\eta}\# = \Omega^{-\frac{1}{2}}\eta\#$, the existence of $\Omega^{-\frac{1}{2}}$ following from the positive definiteness of Ω. Now from (5.22) and (5.19), $\eta\# = XG\#(\Psi\beta - \alpha)$. Under the alternative hypothesis $\Psi_1\beta - \alpha_1 \neq 0$ so $\Psi\beta - \alpha \neq 0$ and consequently $\eta\# \neq 0$ and $\lambda > 0$. Under the null hypothesis $\eta\# = 0$ so $h\#'\Omega^{-1}h\# \sim \sigma^2\chi^2(r_1)$.

We conclude that $h\#'\Omega^{-1}h\#$ and $\varepsilon^{0\prime}\Omega^{-1}\varepsilon^0$ are independently distributed as σ^2 times variables with noncentral and central chi-square distributions respectively. It follows that

$$(5.27) \quad \frac{n - k + r_0}{r_1} \cdot \frac{h\#'\Omega^{-1}h\#}{\varepsilon^{0\prime}\Omega^{-1}\varepsilon^0} = \frac{n - k + r_0}{r_1} \cdot \frac{h'\Omega^{-1}h - h^{0\prime}\Omega^{-1}h^0}{e'\Omega^{-1}e + h^{0\prime}\Omega^{-1}h^0}$$

has the F-distribution with r_1 and $n - k + r_0$ degrees of freedom. We must now show that this is equivalent to the expression in (5.1). From (5.10),

$$h'\Omega^{-1}h = (b'\Psi' - \alpha')G'X'\Omega^{-1}XG(\Psi b - \alpha)$$

(5.28) $$h^{0\prime}\Omega^{-1}h^0 = (b'\Psi' - \alpha')G^{0\prime}X'\Omega^{-1}XG^0(\Psi b - \alpha)$$

$$h\#'\Omega^{-1}h\# = (b'\Psi' - \alpha')G\#'X'\Omega^{-1}XG\#(\Psi b - \alpha).$$

From (5.6) and (5.7),

$$G'X'\Omega^{-1}XG = S^{-1}SS^{-1} = S^{-1}$$

(5.29) $$G^{0\prime}X'\Omega^{-1}XG^0 = ZSZ = Z$$

$$G\#X'\Omega^{-1}XG\# = WSW = W = S^{-1} - Z$$

so

$$h'\Omega^{-1}h = (b'\Psi' - \alpha')S^{-1}(\Psi b - \alpha)$$

(5.30) $$h^{0\prime}\Omega^{-1}h^0 = (b'\Psi' - \alpha')Z(\Psi b - \alpha) = (b'\Psi'_0 - \alpha_0')S_{00}^{-1}(\Psi_0 b - \alpha_0)$$

$$h\#'\Omega^{-1}h\# = (b'\Psi' - \alpha')W(\Psi b - \alpha),$$

and (5.1) is established.

From a computational point of view, expression (5.1) may not be as convenient as desirable, since it requires inversion of the $r \times r$ matrix S, as well as the $r_0 \times r_0$ matrix S_{00}. As the following theorem shows, in addition to S_{00} it is necessary only to invert a matrix of order r_1 rather than $r = r_1 + r_0$:

Theorem 5.2 The F-ratio (5.10) may be written

(5.31) $$F_{r_1,\, n-k+r_0} = \frac{n - k + r_0}{r_1} \cdot \frac{(c_1' - c_0'S_{00}^{-1}S_{01})[S_{11} - S_{10}S_{00}^{-1}S_{01}]^{-1}(c_1 - S_{10}S_{00}^{-1}c_0)}{e'\Omega^{-1}e + c_0'S_{00}^{-1}c_0}$$

where the symbols are defined in (5.2) and (5.3).

Proof. From (5.27) and (5.30), the quadratic form in the numerator of (5.1) is given by

(5.32) $$h\#'\Omega^{-1}h\# = (b'\Psi' - \alpha')W(\Psi b - \alpha) = (b'\Psi' - \alpha')[S^{-1} - Z](\Psi b - \alpha)$$

Now by the rule for inverting block matrices (cf. Aitken[1, p. 139]),

$$(5.33)\quad S^{-1} = \begin{bmatrix} S_{11} & S_{10} \\ S_{01} & S_{00} \end{bmatrix}^{-1} = \begin{bmatrix} (S_{11} - S_{10}S_{00}^{-1}S_{01})^{-1} & -(S_{11} - S_{10}S_{00}^{-1}S_{01})^{-1}S_{10}S_{00}^{-1} \\ -(S_{00} - S_{01}S_{11}^{-1}S_{10})^{-1}S_{01}S_{11}^{-1} & (S_{00} - S_{01}S_{11}^{-1}S_{10})^{-1} \end{bmatrix}.$$

Since S is symmetric,

$$(5.34)\quad (S_{00} - S_{01}S_{11}^{-1}S_{10})^{-1}S_{01}S_{11}^{-1} = S_{00}^{-1}S_{01}(S_{11} - S_{10}S_{00}^{-1}S_{01})^{-1}.$$

Now as we wish to evaluate $W = S^{-1} - Z$ in (5.32), we evaluate

$$\begin{aligned}(5.35)\quad (S_{00} - S_{01}S_{11}^{-1}S_{10})^{-1} - S_{00}^{-1} &= (S_{00} - S_{01}S_{11}^{-1}S_{10})^{-1}[I - (S_{00} - S_{01}S_{11}^{-1}S_{10})S_{00}^{-1}] \\ &= (S_{00} - S_{01}S_{11}^{-1}S_{10})^{-1}S_{01}S_{11}^{-1}S_{10}S_{00}^{-1} \\ &= S_{00}^{-1}S_{01}(S_{11} - S_{10}S_{00}^{-1}S_{01})^{-1}S_{10}S_{00}^{-1},\end{aligned}$$

the last equality being obtained by postmultiplying both sides of (5.34) by $S_{10}S_{00}^{-1}$. Thus we have

$$\begin{aligned}(5.36)\quad W &= S^{-1} - Z \\ &= \begin{bmatrix} (S_{11} - S_{10}S_{00}^{-1}S_{01})^{-1} & -(S_{11} - S_{10}S_{00}^{-1}S_{01})^{-1}S_{10}S_{00}^{-1} \\ -S_{00}^{-1}S_{01}(S_{11} - S_{10}S_{00}^{-1}S_{01})^{-1} & S_{00}^{-1}S_{01}(S_{11} - S_{10}S_{00}^{-1}S_{01})^{-1}S_{10}S_{00}^{-1} \end{bmatrix} \\ &= \begin{bmatrix} I \\ -S_{00}^{-1}S_{01} \end{bmatrix}[S_{11} - S_{10}S_{00}^{-1}S_{01}]^{-1}[I - S_{10}S_{00}^{-1}].\end{aligned}$$

Now using (5.3) we immediately have

$$(5.37)\quad h\#'\Omega^{-1}h\# = c'Wc = (c_1' - c_0'S_{00}^{-1}S_{01})[S_{11} - S_{10}S_{00}^{-1}S_{01}]^{-1}(c_1 - S_{10}S_{00}^{-1}c_0)$$

which establishes the theorem.

The conclusion of Theorem 5.2 can also be arrived at directly by using the restriction $\Psi_0\beta = \alpha_0$ to effect a transformation of variables, and then expressing the restriction $\Psi_1\beta = \alpha_1$ in terms of the transformed variables and using Theorem 4. Since this approach has a certain intrinsic interest, especially from the viewpoint of section 6 to follow, we present it in the following theorem.

<u>Theorem</u> 5.3. The F-ratios (5.1) and (5.31) are equivalent to

(5.38)
$$F_{r^*,n-k^*} = \frac{n-k^*}{r^*} \cdot \frac{h^{*\prime}\Omega^{-1}h^*}{e^{*\prime}\Omega^{-1}e^*} = \frac{n-k^*}{r^*} \cdot \frac{z^{*\prime}\Omega^{-1}H^*z^*}{z^{*\prime}\Omega^{-1}E^*z^*}$$
$$= \frac{n-k^*}{r^*} \cdot \frac{c^*S^{*-1}c^*}{e^{*\prime}\Omega^{-1}e^*}$$

where

$$c^* = \Psi^*b^* - \alpha^*$$
$$S^* = \Psi^*(X^{*\prime}\Omega^{-1}X^*)^{-1}\Psi^{*\prime}$$
$$B^* = (X^{*\prime}\Omega^{-1}X^*)^{-1}X^{*\prime}\Omega^{-1}$$

(5.39)
$$G^* = (X^{*\prime}\Omega^{-1}X^*)^{-1}\Psi^{*\prime}\, S^{*-1}$$
$$R^* = G^*\Psi^*$$
$$D^* = X^*(X^{*\prime}\Omega^{-1}X^*)^{-1}X^{*\prime}\Omega^{-1} = I - E^*$$
$$H^* = X^*R^*D^* = X^*(X^{*\prime}\Omega^{-1}X^*)^{-1}\Psi^{*\prime}S^{*-1}\Psi^*(X^{*\prime}\Omega^{-1}X^*)^{-1}X^{*\prime}\Omega^{-1}$$
$$e^* = y^* - X^*b^*$$
$$h^* = X^*G^*(\Psi^*b^* - \alpha^*)$$
$$z^* = y^* - X^*G^*\alpha^*$$

and where

$$X^* = X\Phi_0$$
$$y^* = y - X\tilde{G}_0\alpha_0$$
$$b^* = B^*y^* = (X^{*\prime}\Omega^{-1}X^*)^{-1}X^{*\prime}\Omega^{-1}y^*$$

(5.40)
$$\Psi^* = \Psi_1\Phi_0$$
$$\alpha^* = \alpha_1 - \Psi_1\tilde{G}_0\alpha_0$$
$$k^* = k - r_0$$
$$r^* = r_1\ ,$$

the matrices Φ_0 and $\tilde{G}_0$ being defined by

$$(5.41)\quad T_0^{-1} = \begin{bmatrix} \tilde{F}_0 \\ \Psi_0 \end{bmatrix}^{-1} = [\Phi_0 \quad \tilde{G}_0]$$

where $\tilde{F}_0$ is some $(k - r_0) \times k$ matrix such that T_0 is invertible.

Proof. From (5.41) we have, as in (3.7),

$$(5.42)\quad T_0T_0^{-1} = \begin{bmatrix} \tilde{F}_0 \\ \Psi_0 \end{bmatrix} [\Phi_0 \quad \tilde{G}_0] = \begin{bmatrix} F_0\Phi_0 & F_0\tilde{G}_0 \\ \Psi_0\Phi_0 & \Psi_0\tilde{G}_0 \end{bmatrix} = \begin{bmatrix} I_{l_0} & 0 \\ 0 & I_{r_0} \end{bmatrix}$$

where $l_0 = k - r_0$. From this we obtain the transformation

$$(5.43)\quad T_0\beta = \begin{bmatrix} \tilde{F}_0\beta \\ \Psi_0\beta \end{bmatrix} = \begin{bmatrix} \beta^* \\ \alpha_0 \end{bmatrix} = \tilde{\beta}$$

and the inverse transformation

$$(5.44)\quad T_0^{-1}\tilde{\beta} = \Phi_0\beta^* + \tilde{G}_0\alpha_0 = \beta\,.$$

Now applying the second set of restrictions to β we obtain

$$(5.45)\quad \Psi_1\beta = \Psi_1\Phi_0\beta^* + \Psi_1G_0\alpha_0 = \alpha_1$$

which, upon applying the definitions for Ψ^* and α^* in (5.40), may be written

$$(5.46)\quad \Psi^*\beta^* = \alpha^*.$$

Now Ψ^* is $r_1 \times l_0 = k - r_0$ and of rank r_1, for, since $[\Psi_1' \quad \Psi_0']$ has rank r, we may take $\tilde{F}_0' = [\tilde{F}' \quad \Psi_1']$ in (5.41), as in (3.5), where $\tilde{F}$ is $l \times k$ of rank $l = n - r$. From (5.42) we have

$$(5.47)\quad I_{l_0} = \tilde{F}_0\Phi_0 = \begin{bmatrix} \tilde{F}\Phi_0 \\ \Psi_1\Phi_0 \end{bmatrix} = \begin{bmatrix} \tilde{F}\Phi_0 \\ \Psi^* \end{bmatrix} = \begin{bmatrix} I_l & 0 \\ 0 & I_{r_1} \end{bmatrix}$$

whence $\Psi^* = [0 \ I_{r_1}]$ is of rank r_1.

Now from (5.44) we have $y = X\beta + \varepsilon = X\Phi_0\beta^* + X\tilde{G}_0\alpha_0 + \varepsilon$, so from the definitions of y^* and X^* in (5.40) this may be written

$$(5.48)\quad y^* = X\beta^* + \varepsilon\,.$$

The problem is then reduced to obtaining an estimate for β^* in (5.48) subject to the restriction (5.46). (5.38) is then an immediate consequence of Theorem 4 and the theorem is proved.

We shall now show that the entities h^*, e^*, H^*, E^* are the same as the previously considered entities $h\#$, ε^0, $H\#$, and E_0.

Theorem 5.4. $D^* = D_0$.

Proof. From (5.12), $D_0 = D - H_0$ by definition. We shall show that $D = D^* + H_0$. Now from (5.42) we have $\Psi_0\Phi_0 = 0$, so using the expressions for D^* and H_0 in (5.39) and (5.11) and the fact that $X^* = X\Phi_0$ from (5.40), it is found that $D^*H_0 = 0 = H_0D^*$. Now it is easily verified that D^* and H_0 are idempotent of ranks $k - r_0$ and r_0 respectively, so from Theorem 1.4, $D^* + H_0$ is idempotent of rank $k - r_0 + r_0 = k$. Furthermore D is idempotent of rank k, and it may be verified that $D[D^* + H_0] = D^* + H_0 = [D^* + H_0]D$. Thus D and $D^* + H_0$ fulfill the conditions of Theorem 1.4, Corollary 2, so $D = D^* + H_0$ and therefore $D^* = D - H_0 = D_0$.

From the definitions of E^* and E_0 in (5.39) and (5.11) an immediate consequence is

Corollary 5.4.1. $E^* = E_0$.

Lemma 5.1. $\Psi^*B^* = [I - S_{10}S_{00}^{-1}]\Psi B$.

Proof. From the definitions of B^*, X^* and Ψ^* in (5.39) and (5.40) we have

$$\Psi^*B^* = \Psi_1\Phi_0(\Phi_0'X'\Omega^{-1}X\Phi_0)^{-1}\Phi_0'X'\Omega^{-1}$$

$$= \Psi_1(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}X\Phi_0(\Phi_0'X'\Omega^{-1}X\Phi_0)^{-1}\Phi_0'X'\Omega^{-1}$$

$$= \Psi_1BD^*$$

where $B = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$ from (5.14). Now it is easily verified that $BD = B$ and $BH_0 = G_0\Psi_0B = R_0B$ where G_0 and R_0 are as defined in (5.7) and (5.8), hence, using Theorem 5.4 we obtain

(5.49) $\Psi^*B^* = \Psi_1 B(D - H_0) = \Psi_1(I - R_0)B = [\Psi_1 - S_{10}S_{00}^{-1}\Psi_0]B$

where S_{10} and S_{00} are given by (5.2).

Lemma 5.2. $S^* = S_{11} - S_{10}S_{00}^{-1}S_{01}$.

Proof. Since $B\Omega B' = (X'\Omega^{-1}X)^{-1}$ and $B^*\Omega B^{*\prime} = X^{*\prime}\Omega^{-1}X^*)^{-1}$, we may apply Lemma 5.1 and write S^* from (5.39) as

(5.50) $S^* = \Psi^*B^*\Omega B^{*\prime}\Psi^{*\prime} = [I \quad - S_{10}S_{00}^{-1}]S[I \quad -S_{10}S_{00}^{-1}]'$

from which the lemma is obtained after computation and cancellation.

Theorem 5.5. $H^* = H\#$.

Proof. From the definitions of H^*, B' and S' in (5.39), and applying Theorem 5.4 and Lemmas 5.1 and 5.2, we have

$$\begin{aligned} H^* &= \Omega B^{*\prime}\Psi^{*\prime}S'^{-1}\Psi^*B^* \\ &= \Omega B'\Psi' \begin{bmatrix} I \\ -S_{00}^{-1}S_{01} \end{bmatrix} [S_{11} - S_{10}S_{00}^{-1}S_{01}]^{-1}[I \quad - S_{10}S_{00}^{-1}]\Psi B \\ &= \Omega B'\Psi' W\Psi B \end{aligned}$$

from (5.36). But this is the same as the definition of $H\#$ as given by (5.13).

Lemma 5.3. $X'(X^{*\prime}\Omega^{-1}X')^{-1}\Psi^{*\prime} = X(X'\Omega^{-1}X)^{-1}\Psi'[I \quad - S_{01}S_{00}^{-1}]'$.

Proof. From the definitions of Ψ^* and X^* in (5.40),

$$\begin{aligned} X^*(X^{*\prime}\Omega^{-1}X')^{-1}\Psi^{*\prime} &= X^*(X^{*\prime}\Omega^{-1}X^*)^{-1}\Phi_0{}'X'\Omega^{-1}X(X'\Omega^{-1}X)^{-1}\Psi_1{}' \\ &= D^*X(X'\Omega^{-1}X)^{-1}\Psi_1{}' \ . \end{aligned}$$

Applying Theorem 5.4 and definition (5.2) this becomes

$$(D - H_0)\ X\ (\ X'\Omega^{-1}X)^{-1}\Psi_1{}' = X(X'\Omega^{-1}X)^{-1}[\Psi_1{}' - \Psi_0{}'S_{00}^{-1}S_{01}]$$

which establishes the Lemma. Together with Lemma 5.2 this result gives, from the definition of G^* in (5.39),

$$(5.51)\quad X^*G^* = X(X'\Omega^{-1}X)^{-1}\Psi'\begin{bmatrix} I \\ -S_{00}^{-1}S_{01}\end{bmatrix}[S_{11} - S_{10}S_{00}^{-1}S_{01}]^{-1}$$

Lemma 5.4. $c^* = [I \quad -S_{10}S_{00}^{-1}]c$.

Proof. From the definition of c^* and B^* in (5.39) and of y^* and α^* in (5.40), and applying Lemma 5.1, we obtain

$$\begin{aligned} c^* &= \Psi^*B^*y^* - \alpha^* \\ &= [\Psi_1 - S_{10}S_{00}^{-1}\Psi_0]B(y - X\tilde{G}_0\alpha_0) - (\alpha_1 - \Psi_1\tilde{G}_0\alpha_0) \\ &= [\Psi_1 - S_{10}S_{00}^{-1}\Psi_0](b - \tilde{G}_0\alpha_0) - (\alpha_1 - \Psi_1\tilde{G}_0\alpha_0). \end{aligned}$$

Since $\Psi_0\tilde{G}_0 = I_{r_0}$ from (5.42) this becomes, after cancellation of $\Psi_1\tilde{G}_0\alpha_0$,

$$\Psi^*b^* - \alpha^* = (\Psi_1 b - \alpha_1) - S_{10}S_{00}^{-1}(\Psi_0 b - \alpha_0) = [I \quad -S_{10}S_{00}^{-1}](\Psi b - \alpha) .$$

Theorem 5.6. $e^* = \varepsilon^0$.

Proof. By definition (5.39), and defining $F_0 = F$ in (3.17), we have, using (3.16)

$$e^* = y^* - X^*b^* = y - X\tilde{G}_0\alpha_0 - X\Phi_0F_0(b - \tilde{G}_0\alpha_0).$$

From (3.19) and (3.23) it is clear that $\Phi_0F_0 = L_0 = I - R_0$ where $R_0 = G_0\Psi_0 = G^0\Psi$ as in (5.8). Thus, using the fact that $R_0\tilde{G}_0 = G_0\Psi_0\tilde{G}_0 = G_0$ from (5.42),

$$\begin{aligned} e^* &= y - X\tilde{G}_0\alpha_0 - X(I - R_0)(b - \tilde{G}_0\alpha_0) \\ &= y - X(L_0 b + G_0\alpha_0) = y - X\beta^0 = \varepsilon^0, \end{aligned}$$

from (3.25), (5.4) and (5.9).

Theorem 5.7. $h^* = h^\#$.

Proof. From the definition of h^* in (5.39), and using (5.51), Lemma 5.4, and (5.36), we have

$$
\begin{aligned}
h^* &= X^*G^*(\Psi^*b^* - \alpha^*)\\
&= X(X'\Omega^{-1}X)^{-1}\Psi' \begin{bmatrix} I \\ -S_{00}^{-1}S_{01} \end{bmatrix} [S_{11} - S_{10}S_{00}^{-1}S_{01}]^{-1}[I \quad - S_{10}S_{00}^{-1}](\Psi b - \alpha)\\
&= X(X'\Omega^{-1}X)^{-1}\Psi' W(\Psi b - \alpha)\\
&= XG^{\#}(\Psi b - \alpha) = h^{\#}
\end{aligned}
$$

from (5.7) and (5.10).

<u>Theorem 5.8.</u> If $\tilde{G}_0 = G_0$ then $X^*G^*\Psi^*b^* = XG^{\#}\Psi b$ and $X^*G^*\alpha^* = XG^{\#}\alpha$.

<u>Proof</u>. From (5.51) and Lemma 5.1, and using (5.36),

$$
\begin{aligned}
X^*G^*\Psi^*b^* &= X(X'\Omega^{-1}X)^{-1}\Psi' \begin{bmatrix} I \\ -S_{00}^{-1}S_{01} \end{bmatrix} [S_{11} - S_{10}S_{00}^{-1}S_{01}]^{-1}[I \quad - S_{10}S_{00}^{-1}]\Psi By^*\\
&= X(X'\Omega^{-1}X)^{-1}\Psi' W\Psi B \quad (y - XG^0\alpha)\\
&= X(X'\Omega^{-1}X)^{-1}\Psi' W\Psi By .
\end{aligned}
$$

Since, from (5.5), (5.6), and (5.7),

$$
\begin{aligned}
W\Psi BXG^0 &= W\Psi(X'\Omega^{-1}X)^{-1}\Psi' Z\\
&= (S^{-1} - Z)SZ = 0.
\end{aligned}
$$

Further, since by hypothesis $\tilde{G}_0 = G_0$ so $\tilde{G}_0\alpha_0 = G_0\alpha_0 = G^0\alpha$, we have $\alpha^* =$ $[I \quad - S_{10}S_{00}^{-1}]\alpha$ so

$$
X^*G^*\alpha^* = X(X'\Omega^{-1}X)^{-1}\Psi' Wy = XG^{\#}\alpha .
$$

We may note that substitution of the expressions for c^* and S^* given by Lemmas 5.4 and 5.2 in the numerator of the last expression in (5.38) gives directly the numerator of (5.31). Similarly from Corollary 5.4.1 and (5.12), and since $EX = 0$

$$
e^* = E^*y^* = (E + H_0)(y - X\tilde{G}_0\alpha_0) = Ey + XG_0(\Psi_0 b - \alpha_0)
$$

and since $E'\Omega^{-1}X = 0$,

$$
e^*\Omega^{-1}e^* = y'E'\Omega^{-1}Ey + (b'\Psi_0' - \alpha_0)G_0X'\Omega^{-1}XG_0'(\Psi_0 b - \alpha_0) = e'\Omega^{-1}e + c_0'S_{00}^{-1}c_0
$$

giving directly the denominator of (5.31) from the denominator of the last expression in (5.38).

It is of interest to note that (5.31) may be written in a slightly different form. From section 2 we know that the unrestricted Markov estimate of $\Psi_0\beta$ is given by

$$(5.52) \quad a_0 = \Psi_0 b$$

and the Markov estimate of $\Psi_1\beta$ subject to the restriction $\Psi_0\beta = \alpha_0$ is given by

$$(5.53) \quad a_1 = \Psi_1\beta^0 = [\Psi_1 - S_{10}S_{00}^{-1}\Psi_0]b + S_{10}S_{00}^{-1}\alpha_0.$$

Being Markov estimates they are of course unbiased, so $\mathcal{E}a_0 = \alpha_0$ and $\mathcal{E}a_1 = \alpha_1$. So from (5.52 we have

$$(5.54) \quad c_0 = a_0 - \alpha_0$$

where $c_0 = \Psi_0 b - \alpha_0$ from (5.3), and from Lemma 5.4 and (5.53) we have

$$(5.55) \quad c^* = (\Psi_1 b - \alpha_1) - S_{10}S_{00}^{-1}(\Psi_0 b - \alpha_0)$$

$$= [\Psi_1 - S_{10}S_{00}^{-1}\Psi_0]b + S_{10}S_{00}^{-1}\alpha_0 - \alpha_1$$

$$= a_1 - \alpha_1.$$

Thus c^*_0 and c^* are simply the deviations of the Markov estimates a_0 and a_1 from their means. Furthermore $\mathcal{E}e = 0$, so the statistics e, a_0 and a_1 correspond to the situations of no restrictions, one set of restrictions, and both sets of restrictions, respectively. (5.31) may then be written

$$(5.56) \quad F_{r_1, n-k+r_0} = \frac{n-k+r_0}{r_1} \cdot \frac{(a_1' - \alpha_1')[S_{11} - S_{10}S_{00}^{-1}S_{01}]^{-1}(a_1 - \alpha_1)}{e'\Omega^{-1}e + (a_0' - \alpha_0')S_{00}^{-1}(a_0 - \alpha_0)}.$$

If we allow for the existence of null or empty matrices, the results of section 4 are special cases of Theorem 5.31. If Ψ_0 is the empty matrix (denoted $\Psi_0 = \emptyset$) with k columns and no rows, then it is trivially of full rank (zero) and the matrices $\tilde{F}_0$ and Φ_0 in (5.42) are $k \times k$ of rank k. Using the convention that the product of an empty $0 \times k$ matrix and a $k \times i$ matrix is the empty $0 \times i$ matrix, and that the inverse of the empty 0×0 matrix (which is of full rank and is also denoted by ϕ) is itself the empty 0×0 matrix, then if $\Psi_0 = \phi$ we have $a_0 = \phi$, $S_{00} = \phi$, $S_{00}^{-1} = \phi$, $a_1 = \Psi_1 b - \alpha_1$, and (5.31) is equivalent to (4.1) except that Ψ_1 takes the place of Ψ.

6. Linear Restrictions and Projections in Linear Spaces.

6.0. Preliminaries

Because of the equivalence of idempotent linear transformations and projections in linear (vector) spaces, the results of the previous sections have a natural geometric interpretation, which will be developed in the present section. The mathematical concepts underlying this section may be found in Halmos [15, especially pp.71-78] and Jacobson [18, esp. Vol. II, Ch. IV]. For convenience we summarize here some of the main definitions and results.

Definition 6.1. A subspace (or linear manifold) $\mathcal{E}$ of a linear space $\mathcal{Y}$ is a subset of $\mathcal{Y}$ such that $x, y \in \mathcal{E}$ implies that $\kappa x + \lambda y \in \mathcal{E}$, where κ and λ are scalars.

Definition 6.2. A coset (or linear variety) is a set of vectors $z = x + y$ where x is fixed and $y \in \mathcal{E}$, $\mathcal{E}$ being a subspace. It is denoted $\mathcal{E} + y$, where the symbol y refers in this context to the set consisting of the single vector y.

Definition 6.3. If $\mathcal{A}$ and $\mathcal{B}$ are two subspaces of a linear space $\mathcal{Y}$ such that every $z \in \mathcal{Y}$ may be written uniquely as $z = x + y$, with $x \in \mathcal{A}$ and $y \in \mathcal{B}$, then $\mathcal{Y}$ is called the direct sum of $\mathcal{A}$ and $\mathcal{B}$, and we write $\mathcal{Y} = \mathcal{A} \oplus \mathcal{B}$. Equivalently, if $\mathcal{A} + \mathcal{B} = \{x + y \mid x \in \mathcal{A}, y \in \mathcal{B}\} = \mathcal{Y}$ and $\mathcal{A} \cap \mathcal{B} = \{z \mid z \in \mathcal{A}, z \in \mathcal{B}\} = 0$, where 0 denotes the set consisting of the zero vector or origin 0, then $\mathcal{Y} = \mathcal{A} \oplus \mathcal{B}$. We will occasionally denote $\mathcal{B} = \bar{\mathcal{A}}$.

Definition 6.4. If $\mathcal{E}$ is a subspace of $\mathcal{Y}$, and A a linear transformation, then the set $A\mathcal{E} = \{y \mid y = Ax, x \in \mathcal{E}\}$ is called the image of $\mathcal{E}$ under A. $\mathcal{A} = A\mathcal{Y}$ is called the range of A. The set $\{x \mid Ax = 0\}$ is called the null space of A.

Definition 6.5. A subspace $\mathcal{E}$ is said to be invariant under a linear transformation A if $A\mathcal{E} \subseteq \mathcal{E}$ (i.e., if $x \in \mathcal{E} \rightarrow Ax \in \mathcal{E}$), and $\mathcal{E}$ is said to be reduced by A.

Defintion 6.6. If $\mathcal{Y} = \mathcal{E} \oplus \bar{\mathcal{E}}$ and $z = x + y$ where $x \in \mathcal{E}$ and $y \in \bar{\mathcal{E}}$, then the linear transformation defined by $Ez = x$ is called the projection of $\mathcal{Y}$ on $\mathcal{E}$ along $\bar{\mathcal{E}}$, which we denote $E: \mathcal{Y} \to \mathcal{E} \mid \bar{\mathcal{E}}$.

Result 6.1. A linear transformation E is a projection if and only if it is idempotent.

Result 6.2. If E is a linear transformation, $E: \mathcal{Y} \to \mathcal{E} \mid \bar{\mathcal{E}}$ if and only if $I - E: \mathcal{Y} \to \bar{\mathcal{E}} \mid \mathcal{E}$, where $\mathcal{E} = E\mathcal{Y}$, $\bar{\mathcal{E}} = (I - E)\mathcal{Y}$, and $\mathcal{E} \oplus \bar{\mathcal{E}} = \mathcal{Y}$. I.e., the range and null space of a projection are complementary subspaces whose direct sum is the whole space.

Result 6.3. Let $A: \mathcal{Y} \to \mathcal{A} \mid \bar{\mathcal{A}}$, $B: \mathcal{Y} \to \mathcal{B} \mid \bar{\mathcal{B}}$.

Then

(i) If and only if AB = BA = 0,

$A + B : \mathcal{Y} \to \mathcal{A} \oplus \mathcal{B} \mid \bar{\mathcal{A}} \cap \bar{\mathcal{B}}$

(ii) If and only if AB = BA = B,

$A - B : \mathcal{Y} \to \mathcal{A} \cap \bar{\mathcal{B}} \mid \bar{\mathcal{A}} \oplus \mathcal{B}$

(iii) If AB = BA,

$AB : \mathcal{Y} \to \mathcal{A} \cap \mathcal{B} \mid \bar{\mathcal{A}} + \bar{\mathcal{B}}$

In view of Result 6.2, we have in these three cases

(i) $(A + B)\mathcal{Y} = \mathcal{A} \oplus \mathcal{B}$; $(I - A - B)\mathcal{Y} = \bar{\mathcal{A}} \cap \bar{\mathcal{B}}$.

(ii) $(A - B)\mathcal{Y} = \mathcal{A} \cap \bar{\mathcal{B}}$; $(I - A = B)\mathcal{Y} = \bar{\mathcal{A}} \oplus \mathcal{B}$

(iii) $AB\mathcal{Y} = \mathcal{A} \cap \mathcal{B}$; $(I - AB)\mathcal{Y} = \bar{\mathcal{A}} + \bar{\mathcal{B}}$.

Result 6.4. If $\mathcal{Y} = \mathcal{E} \oplus \bar{\mathcal{E}}$, then $\mathcal{E}$ and $\bar{\mathcal{E}}$ are invariant under (reduced by) the linear transformation A (i.e., $A\mathcal{E} \subseteq \mathcal{E}$ and $A\bar{\mathcal{E}} \subseteq \bar{\mathcal{E}}$) if and only if EA = AE, where $E: \mathcal{Y} \to \mathcal{E} \mid \bar{\mathcal{E}}$ and $\mathcal{E} = E\mathcal{Y}$.

Lemma 6.1. If AB = B and BA = A, then $\mathcal{A} = \mathcal{B}$ where $\mathcal{A} = A\mathcal{Y}$ and $\mathcal{B} = B\mathcal{Y}$.

Proof. From Definition 6.4 it is clear that $A\mathcal{B} \subseteq A\mathcal{Y} = \mathcal{A}$ and $B\mathcal{A} \subseteq B\mathcal{Y} = \mathcal{B}$. Since AB = B and BA = A we also have $A\mathcal{B} = AB\mathcal{Y} = B\mathcal{Y} = \mathcal{B}$ and $B\mathcal{A} = BA\mathcal{Y} = A\mathcal{Y} = \mathcal{A}$, hence $\mathcal{A} = B\mathcal{A} \subseteq \mathcal{B}$ and $\mathcal{B} = A\mathcal{B} \subseteq \mathcal{A}$, so $\mathcal{A} = \mathcal{B}$.

6.1. Homomorphism of Invariant Subspaces in the Parameter and Sample Spaces.

In sections 3-5 we have come across $k \times k$ idempotent matrices R, R_0, $R\#$, operating in the k-dimensional parameter space. We repeat here the definitions of these matrices, together with their complements, where $\Psi' = [\Psi'_1 \quad \Psi'_0]$:

$$
\begin{aligned}
& R = (X'\Omega^{-1}X)^{-1}\Psi'[\Psi(X'\Omega^{-1}X)^{-1}\Psi']^{-1} = I - L \\
(6.1) \quad & R_0 = (X'\Omega^{-1}X)^{-1}\Psi_0'[\Psi_0(X'\Omega^{-1}X)^{-1}\Psi_0']^{-1}\Psi_0 = I - L_0 \\
& R\# = R - R_0 = I - L\# = L_0 - L.
\end{aligned}
$$

From (5.17), $RR_0 = R_0 = R_0R$, whence $LL_0 = L = L_0L$, and $RR\# = R\# = R\#R$, etc.

Similarly we have come across the $n \times n$ idempotent matrices H, H_0, $H\#$. We define these, together with their complements, in terms of the above, as

$$
\begin{aligned}
& H = XRB = I - J \\
(6.2) \quad & H_0 = XR_0B = I - J_0 \\
& H\# = H - H_0 = I - J\# = J_0 - J
\end{aligned}
$$

where $B = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$. From (5.18), $HH_0 = H_0 = H_0H$, whence $JJ_0 = J = J_0J$, and $JJ\# = J = J\#J$. It is natural to introduce the following:

Definition 6.7. Two square matrices U and V of orders n and k respectively, where $n \geqq k$, will be said to be quasi-similar if there exists an $n \times k$ matrix X of rank k such that $UX = XV$.

Lemma 6.2. Let $\mathcal{S}_n$ be a set of $n \times n$ matrices U, and $\mathcal{S}_k$ a set of $k \times k$ matrices V, and X an $n \times k$ matrix of rank k. If for all $V \in \mathcal{S}_k$ there exists a $U \in \mathcal{S}_n$ such that $UX = XV$, then the transformation $UX = XV$ defines a homomorphism of $\mathcal{S}_n$ onto $\mathcal{S}_k$, and $\mathcal{S}_k$ is the homomorphic image of $\mathcal{S}_n$.

Proof. Since X has rank k, there exists an $n \times n$ matrix C such that $X'CX$ is invertible. Define $Y = (X'CX)^{-1}X'C$; then for all such C, $YUX = V$, so for each U there corresponds a unique V. Thus the transformation $UX = XV$ defines a mapping φ of $\mathcal{S}_n$ onto $\mathcal{S}_k$. Now let $U_1, U_2 \in \mathcal{S}_n$, and $\varphi(U_1) = V_1$, $\varphi(U_2) = V_2$. Then $U_1X = XV_1$ and $U_2X = XV_2$, so $(U_1 + U_2)X = X(V_1 + V_2)$ and

$U_1U_2X = U_1XV_2 = XV_1V_2$. Thus $\varphi(U_1 + U_2) = \varphi(U_1) + \varphi(U_2)$ and $\varphi(U_1U_2) = \varphi(U_1) \cdot \varphi(U_2)$.

We may note immediately that if $M = XY = X(X'CX)^{-1}X'C$, then $MX = XI_k$; thus I_n and matrices such as M have I_k as their homomorphic image, and consequently M plays the role of a multiplicative identity element. This collection of matrices I_n, M, ... is called the (multiplicative) kernel of the homomorphism.

Just as I_n and $D = X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1}$ form the multiplicative kernel of the homomorphism between the projections in the sample and parameter spaces, 0_n and $E = I - D$ form the additive kernel, since their homomorphic image under φ is 0_k. From the definitions

$$D = X(X'\Omega^{-1}X)^{-1}X'\Omega^{-1} = I - E$$

$$D_0 = I - E_0 = I - (E + H_0) = D - H_0$$

$$D\# = I - E\# = I - (E + H\#) = D - H\# \tag{6.3}$$

$$D = I - E = I - (E + H) = D - H = D_0 - H\# = I - (E_0 + H\#)$$

we may display the homomorphism as follows, where on the left margin is shown, for convenience, the rank of the n × n matrix on the left, and on the right side is shown the rank of both the center n × n matrix and the k × k matrix on the right:

(6.4)

n	$I_nX = DX = XI_k$	k
$n - k$	$EX = 0_nX = X0_k$	0
$n - r$	$JX = \hat{D}X = XL$	$l = k - r$
$n - k + r$	$\hat{E}X = HX = XR$	r
$n - r_0$	$JX = D_0X = XL_0$	$l_0 = k - r_0$
$n - k + r_0$	$E_0X = H_0X = XR_0$	r_0
$n - r_1$	$J\#X = D\#X = XL\#$	$l_1 = k - r_1$
$n - k + r_1$	$E\#X = H\#X = XR\#$	r_1

Denoting the ranges of matrices by the corresponding script letters, we have the following projections in the n-dimensional sample space $\mathcal{N}$ and k-dimensional parameter space $\mathcal{K}$:

$$
\begin{array}{lll}
 & D: \mathcal{N} \rightarrow \mathcal{D} \mid \mathcal{E} & I_k: \mathcal{K} \rightarrow \mathcal{K} \mid 0 \\
 & E: \mathcal{N} \rightarrow \mathcal{E} \mid \mathcal{D} & O_k: \mathcal{K} \rightarrow 0 \mid \mathcal{K} \\
 & J: \mathcal{N} \rightarrow \mathcal{J} \mid \mathcal{H} & L: \mathcal{K} \rightarrow \mathcal{L} \mid \mathcal{R} \\
(6.5) & H: \mathcal{N} \rightarrow \mathcal{H} \mid \mathcal{J} & R: \mathcal{K} \rightarrow \mathcal{R} \mid \mathcal{L} \\
 & J_0: \mathcal{N} \rightarrow \mathcal{J}_0 \mid \mathcal{H}_0 & L_0: \mathcal{K} \rightarrow \mathcal{L}_0 \mid \mathcal{R}_0 \\
 & H_0: \mathcal{N} \rightarrow \mathcal{H}_0 \mid \mathcal{J}_0 & R_0: \mathcal{K} \rightarrow \mathcal{R}_0 \mid \mathcal{L}_0
\end{array}
$$

From (4.14), (5.5), (5.17), (5.18) we also have, making use of Result 6.3,

$$
\begin{array}{lll}
 & \hat{D}: \mathcal{N} \rightarrow \hat{\mathcal{D}} = \mathcal{D}_0 \cap \mathcal{J}\# \mid \hat{\mathcal{E}} = \mathcal{E}_0 \oplus \mathcal{H}\# & L: \mathcal{K} \rightarrow \mathcal{L} = \mathcal{L}_0 \cap \mathcal{L}\# \mid \mathcal{R} = \mathcal{R}_0 \oplus \mathcal{R}\# \\
 & \hat{E}: \mathcal{N} \rightarrow \hat{\mathcal{E}} = \mathcal{E}_0 \oplus \mathcal{H}\# \mid \hat{\mathcal{D}} = \mathcal{D}_0 \cap \mathcal{J}\# & R: \mathcal{K} \rightarrow \mathcal{R} = \mathcal{R}_0 \oplus \mathcal{R}\# \mid \mathcal{L} = \mathcal{L}_0 \cap \mathcal{L}\# \\
 & D_0: \mathcal{N} \rightarrow \mathcal{D}_0 = \mathcal{D} \cap \mathcal{J}_0 \mid \mathcal{E}_0 = \mathcal{E} \oplus \mathcal{H}_0 & L_0: \mathcal{K} \rightarrow \mathcal{L}_0 = \mathcal{K} \cap \mathcal{L}_0 \mid \mathcal{R}_0 = 0 \oplus \mathcal{R}_0 \\
(6.6) & E_0: \mathcal{N} \rightarrow \mathcal{E}_0 = \mathcal{E} \oplus \mathcal{H}_0 \mid \mathcal{D}_0 = \mathcal{D} \cap \mathcal{J}_0 & R_0: \mathcal{K} \rightarrow \mathcal{R}_0 = 0 \oplus \mathcal{R}_0 \mid \mathcal{L}_0 = \mathcal{K} \cap \mathcal{L}_0 \\
 & D\#: \mathcal{N} \rightarrow \mathcal{D}\# = \mathcal{D} \cap \mathcal{J}\# \mid \mathcal{E}\# = \mathcal{E} \oplus \mathcal{H}\# & L\#: \mathcal{K} \rightarrow \mathcal{L}\# = \mathcal{K} \cap \mathcal{L}\# \mid \mathcal{R}\# = 0 \oplus \mathcal{R}\# \\
 & E\#: \mathcal{N} \rightarrow \mathcal{E}\# = \mathcal{E} \oplus \mathcal{H}\# \mid \mathcal{D}\# = \mathcal{D} \cap \mathcal{J}\# & R\#: \mathcal{K} \rightarrow \mathcal{R}\# = 0 \oplus \mathcal{R}\# \mid \mathcal{L}\# = \mathcal{K} \cap \mathcal{L}\# \\
 & J\#: \mathcal{N} \rightarrow \mathcal{J}\# = \mathcal{J} \oplus \mathcal{H}_0 \mid \mathcal{K}\# = \mathcal{H} \cap \mathcal{J}_0 & L\#: \mathcal{K} \rightarrow \mathcal{L}\# = \mathcal{L} \oplus \mathcal{R}_0 \mid \mathcal{R}\# = \mathcal{R} \cap \mathcal{L}_0 \\
 & H\#: \mathcal{N} \rightarrow \mathcal{H}\# = \mathcal{H} \cap \mathcal{J}_0 \mid \mathcal{J}\# = \mathcal{J} \oplus \mathcal{H}_0 & R\#: \mathcal{K} \rightarrow \mathcal{R}\# = \mathcal{R} \cap \mathcal{L}_0 \mid \mathcal{L}\# = \mathcal{L} \oplus \mathcal{R}_0
\end{array}
$$

Now it is easily verified, making use of (5.15) and (5.18), that the $n \times n$ matrices in (6.4) are commutative as well as idempotent. Consequently they form a commutative semi-group under multiplication; and from result 6.4 the range and null space of each matrix is invariant with respect to all matrices in the semi-group. Likewise the $k \times k$ matrices in (6.4) form a commutative semi-group under multiplication, and their ranges are also invariant subspaces with respect to the elements of the semi-group. Suitably ordered, the set of invariant subspaces in both the sample and parameter space forms a complemented modular lattice, and these two lattices are homomorphic.

Of particular interest in the restriction of the projections in the sample space to the subspace $\mathcal{D}$. It will be noted that the $n \times n$ matrices in the center of (6.4) are obtained by either premultiplying or postmultiplying the corresponding matrix shown on the left, by D. We introduce

Definition 6.8. Two square matrices U and V of orders n and k respectively, where $n \geq k$, will be said to be similar if there exist $n \times k$ matrices X and Y', with the property $YX = I_k$, such that $U = XVY$.

Lemma 6.3. Let $\mathcal{S}_n$ and $\mathcal{S}_k$ be sets of $n \times n$ and $k \times k$ matrices respectively, where $n \geq k$, connected by the similarity transformation $U = XVY$ where $U \in \mathcal{S}_n$ and $V \in \mathcal{S}_k$. Then this transformation is an isomorphism between $\mathcal{S}_n$ and $\mathcal{S}_k$, with kernels $O_n = XO_kY$ and $M = XI_kY$.

Proof. Since postmultiplication by X gives $UX = XV$, the similarity transformation is a homomorphism of $\mathcal{S}_n$ onto $\mathcal{S}_k$ by Lemma 6.2. Since $U = XVY$ it is also one-to-one, hence an isomorphism.

The isomorphism in question is shown by the relations in (6.4) between the matrices in the center and right column. Restricted to $\mathcal{D}$, the projections in the sample space are as follows:

$$
(6.7)\qquad
\begin{array}{ll}
D: \mathcal{D} \to \mathcal{D} \mid 0 & O_n: \mathcal{D} \to 0 \mid \mathcal{D} \\
J: \mathcal{D} \to \hat{\mathcal{D}} \mid \mathcal{H} & H: \mathcal{D} \to \mathcal{H} \mid \hat{\mathcal{D}} \\
J_0: \mathcal{D} \to \mathcal{D}_0 \mid \mathcal{H}_0 & H_0: \mathcal{D} \to \mathcal{H}_0 \mid \mathcal{D}_0 \\
J\#: \mathcal{D} \to \mathcal{D}\# \mid \mathcal{H}\# & H\#: \mathcal{D} \to \mathcal{H}\# \mid \mathcal{D}\#
\end{array}
$$

This establishes the isomorphism between invariant subspaces in the subspace $\mathcal{D}$ of the sample space, and in the parameter space $\mathcal{K}$, as shown in the following table:

Sample Space	Parameter Space
$\mathcal{D} = \hat{\mathcal{D}} \oplus \mathcal{H}$	$\mathcal{K} = \mathcal{L} \oplus \mathcal{R}$
$\hat{\mathcal{D}} = \mathcal{D}_0 \cap \mathcal{D}\#$	$\mathcal{L} = \mathcal{L}_0 \cap \mathcal{L}\#$
$\mathcal{H} = \mathcal{H}_0 \oplus \mathcal{H}\#$	$\mathcal{R} = \mathcal{R}_0 \oplus \mathcal{R}\#$

In considering the homomorphism and isomorphism discussed above, a special case is of particular interest. In practical applications one will usually have

$\Omega = I$, and in certain experiments and models one will also have $X' = [I_k \quad 0]$, where 0 is $k \times (n - k)$. Then for the matrices U_1 and U_2 in the left and center columns of (6.4) one will have, respectively

$$U_1 = \begin{bmatrix} V & 0 \\ 0 & I_{n-k} \end{bmatrix} \quad U_2 = \begin{bmatrix} V & 0 \\ 0 & 0_{n-k} \end{bmatrix}$$

where V is the corresponding $k \times k$ matrix on the right. In this case there is a congruent embedding of the parameter space in the sample space.

6.2. Projections in the Parameter Space

The matrices $\tilde{L} = \Phi\tilde{F}$ and $\tilde{R} = \tilde{G}\Psi$ of (3.8) were seen to be $k \times k$ idempotent matrices of ranks l and r respectively, with $\tilde{L} + \tilde{R} = I_k$ and $l + r = k$. Thus they are projections on their ranges $\tilde{\mathcal{L}} = \tilde{L}\mathcal{K}$ and $\tilde{\mathcal{R}} = \tilde{R}\mathcal{K}$ where $\mathcal{K} = \mathcal{I}_k$ is the k-dimensional parameter space and $\tilde{\mathcal{L}} \oplus \tilde{\mathcal{R}} = \mathcal{K}$ (i.e., $\tilde{L} : \mathcal{K} \to \tilde{\mathcal{L}} \mid \tilde{\mathcal{R}}$ and $\tilde{R} : \mathcal{K} \to \tilde{\mathcal{R}} \mid \tilde{\mathcal{L}}$). Any vector $\beta \in \mathcal{K}$ may be expressed uniquely in the form $\beta = \lambda + \rho$ where $\lambda \in \tilde{\mathcal{L}}$ and $\rho \in \tilde{\mathcal{R}}$; this is expressed by (3.11) where $\lambda = \Phi\beta^*$ and $\rho = \tilde{G}\alpha$. To see that $\Phi\beta^* \in \tilde{\mathcal{L}}$ we note that it is equal to its own projection on $\tilde{\mathcal{L}}$ along $\tilde{\mathcal{R}}$, that is, $\tilde{L}\lambda = \Phi\tilde{F}\Phi\beta^* = \Phi\beta^* = \lambda$, since $\tilde{F}\Phi = I_l$ from (3.7). Similarly $\tilde{G}\alpha \in \tilde{\mathcal{R}}$ since $R\rho = \tilde{G}\Psi\tilde{G}\alpha = \tilde{G}\alpha = \rho$, where $\Psi\tilde{G} = I_r$ from (3.7). The effect of the restriction (3.2) is to confine the admissible β to the coset $\mathcal{K}^* = \tilde{\mathcal{L}} + \rho$ where $\rho = \tilde{G}\alpha$.

In similar fashion $\mathcal{K} = \mathcal{I}_k$ may be considered as the direct sum of the ranges $\mathcal{I}_l$ and $\mathcal{I}_r$ of $\tilde{F}\Phi = I_l$ and $\Psi\tilde{G} = I_r$ respectively, i.e., $\mathcal{K} = \mathcal{I}_k = \mathcal{I}_l \oplus \mathcal{I}_r$, and (3.7) may be written $I_k = I_l \oplus I_r$. From (3.9), any vector $\tilde{\beta} \in \mathcal{K}$ may be written

$$\tilde{\beta} = \begin{bmatrix} \tilde{F}\beta \\ \Psi\beta \end{bmatrix}$$

where $\tilde{F}\beta \in \mathcal{I}_l$ and $\Psi\beta \in \mathcal{I}_r$, and the restriction (3.2) fixes $\Psi\beta = \alpha$. The subspaces $\tilde{\mathcal{L}}$ and $\tilde{\mathcal{R}}$ are respectively isomorphic to the subspaces $\mathcal{I}_l$ and $\mathcal{I}_r$, but they are not congruent in general. However, in the special case in which Ψ and $\tilde{F}$ consist of coordinate vectors only, we have $\tilde{G} = \Psi'$ and $\Phi = \tilde{F}'$, so that

$$\tilde{L} = \begin{bmatrix} I_1 & 0 \\ & \\ 0 & 0 \end{bmatrix} \qquad \tilde{R} = \begin{bmatrix} 0 & 0 \\ & \\ 0 & I_r \end{bmatrix}$$

In Figure 1 an illustration is given for the case

$$X = \begin{bmatrix} 1 & 1 \\ 1 & 0 \\ 1 & -2 \end{bmatrix} \qquad y = \begin{bmatrix} 4\frac{1}{2} \\ -2 \\ -1 \end{bmatrix} \qquad \begin{matrix} \Psi = (-1 \quad 1) \\ \\ \tilde{F} = (1 \quad 0) \end{matrix} \qquad \begin{matrix} \alpha = 1 \\ \\ \Omega = I \end{matrix}$$

The subspaces $\tilde{\mathcal{L}}$ and $\tilde{\mathcal{R}}$ are straight lines through the origin of the Euclidean plane $\mathcal{K}$, $\tilde{\mathcal{R}}$ being taken (by choice of F) as the vertical axis. The coset $\mathcal{K}^* = \tilde{\mathcal{L}} + \rho$ is the straight line parallel to $\tilde{\mathcal{L}}$ and passing through $\rho = \tilde{G}\alpha = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$; it has the analytic representation $\mathcal{K}^* = \{\beta \mid \Psi\beta = \alpha\}$ or, in this case, $\beta_2 - \beta_1 = 1$.

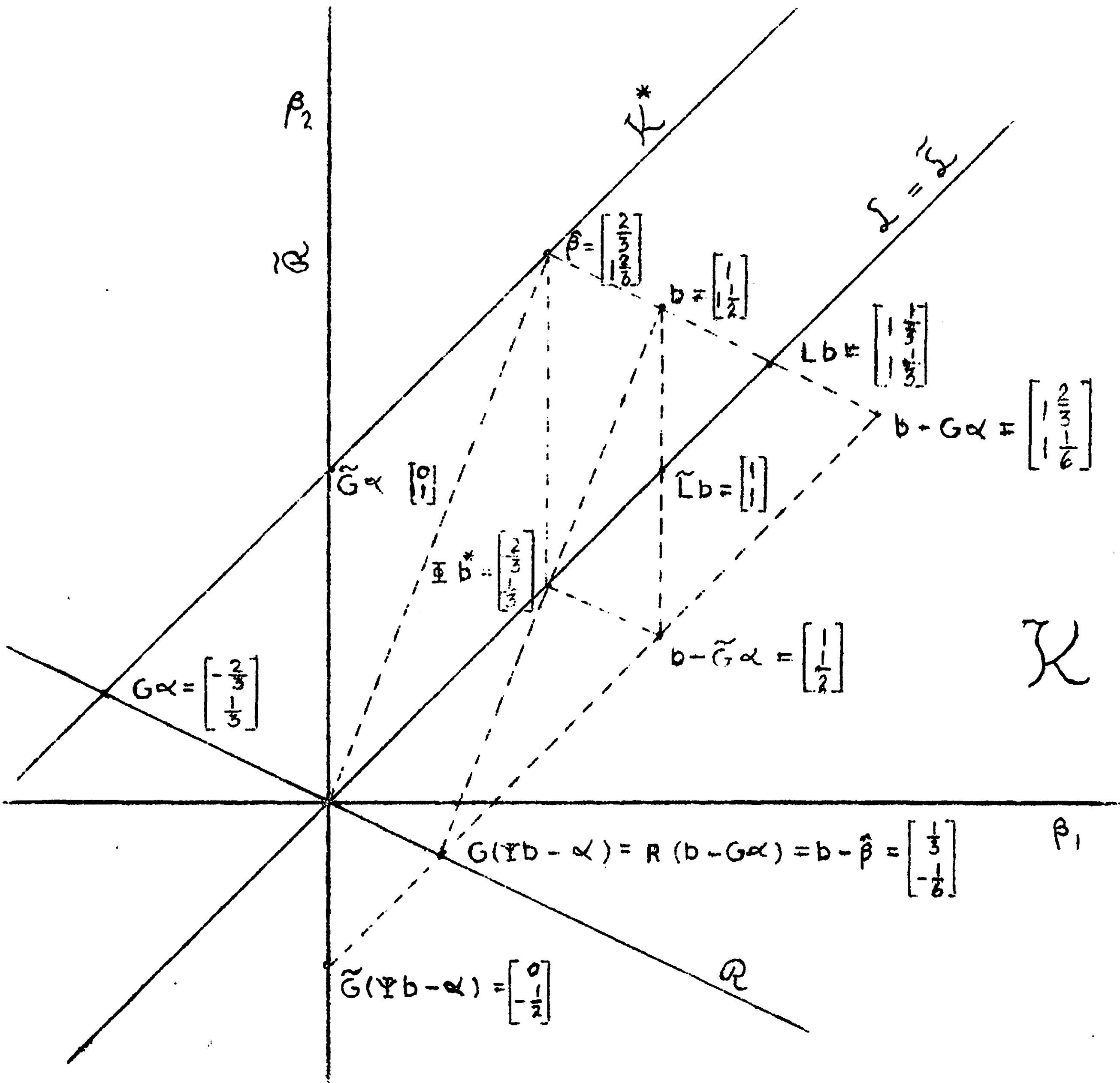
Figure 1 also shows the subspaces $\mathcal{L} = L\mathcal{K}$ and $\mathcal{R} = R\mathcal{K}$ where $L = \Phi F$ and $R = G\Psi$ from (3.19) and (3.22). Since $LL = L$, $RR = R$, $LR = RL = 0$, and $L + R = I_k$, we have $L: \mathcal{K} \to \mathcal{L} \mid \mathcal{R}$ and $R: \mathcal{K} \to \mathcal{R} \mid \mathcal{L}$. Note further that $\tilde{F}\Phi = I_k$ from (3.7) and $F\Phi = I_k$ from (3.17), so from the definitions $\tilde{L} = \Phi\tilde{F}$ in (3.8) and $L = \Phi F$ in (3.19) we have

$$(6.8) \qquad \tilde{L}L = L \; ; \quad L\tilde{L} = \tilde{L}$$

for all L, hence for all X of required rank. Applying Lemma 6.1 to (6.8) we immediately have

$$(6.9) \qquad \tilde{\mathcal{L}} = \mathcal{L}$$

for all X. (6.9) has two interpretation: first, $\mathcal{L}$ is invariant with respect to choice of $\tilde{F}$, so we could of course have chosen $\tilde{F} = F$; second, $\mathcal{L}$ is invariant with respect to X, and so does not depend on X in any way except, of course, its rank. Another way of expressing (6.9) is to say that $\mathcal{L}$ depends on the model and not on the experiment or observations.



β_2
β_1
$\hat{\beta} = \begin{bmatrix} \frac{2}{3} \\ 1\frac{2}{6} \end{bmatrix}$
$b = \begin{bmatrix} 1 \\ 1\frac{1}{2} \end{bmatrix}$
$Lb = \begin{bmatrix} 1\frac{1}{3} \\ 1\frac{1}{3} \end{bmatrix}$
$b - G\alpha = \begin{bmatrix} 1\frac{2}{3} \\ 1\frac{1}{6} \end{bmatrix}$
$\tilde{G}\alpha \begin{bmatrix} 0 \\ 1 \end{bmatrix}$
$\tilde{L}b = \begin{bmatrix} 1 \\ 1 \end{bmatrix}$
$\Phi b^* = \begin{bmatrix} \frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$
$b - \tilde{G}\alpha = \begin{bmatrix} 1 \\ \frac{1}{2} \end{bmatrix}$
$G\alpha = \begin{bmatrix} -\frac{2}{3} \\ \frac{1}{3} \end{bmatrix}$
$G(\Psi b - \alpha) = R(b - G\alpha) = b - \hat{\beta} = \begin{bmatrix} \frac{1}{3} \\ -\frac{1}{6} \end{bmatrix}$
$\tilde{G}(\Psi b - \alpha) = \begin{bmatrix} 0 \\ -\frac{1}{2} \end{bmatrix}$
$\mathcal{L}^*$
$\mathcal{L} = \tilde{\mathcal{L}}$
$\mathcal{K}$
$\mathcal{R}$

Figure 1

On the other hand $\mathcal{R}$ does depend on the observations X, since from the definitions $\tilde{R} = \tilde{G}\Psi$ and $R = G\Psi$ we have $\tilde{R}R = \tilde{G}\Psi G\Psi = \tilde{G}\Psi = \tilde{R}$ and similarly $R\tilde{R} = R$, from which we can conclude only that $\tilde{R}\mathcal{R} = \tilde{\mathcal{R}} = \tilde{R}\mathcal{K}$ and $R\tilde{\mathcal{R}} = \mathcal{R} = R\mathcal{K}$, i.e., that the images of $\mathcal{R}$ and $\tilde{\mathcal{R}}$ under R and $\tilde{R}$ respectively are equal to the ranges of $\tilde{R}$ and R. Consequently $\mathcal{R} \neq \tilde{\mathcal{R}}$ in general, so $\mathcal{R}$, which depends on X, will in general be different for different X.

But although $\mathcal{R} \neq \tilde{\mathcal{R}}$ in general, nevertheless $\mathcal{K}^* = \tilde{\mathcal{K}}^*$, where $\mathcal{K}^* = \mathcal{L} + G\alpha$ and $\tilde{\mathcal{K}}^* = \mathcal{L} + \tilde{G}\alpha$. This follows from equations (3.18) and (3.23) according to which

$$\hat{\beta} = \Phi b^* + \tilde{G}\alpha = Lb + G\alpha \tag{6.10}$$

where Φ and $\tilde{G}$ depend on $\tilde{F}$. $\Phi b^* \in \mathcal{L} = \tilde{\mathcal{L}}$ since it is its own projection on $\tilde{\mathcal{L}}$ along $\tilde{\mathcal{R}}$, i.e., $\tilde{L}\Phi b^* = \Phi F\Phi b^* = \Phi b^*$ since $F\Phi = I_1$ from (3.7). Clearly $Lb \in \mathcal{L}$ also. So from (6.10), $\mathcal{L} + \tilde{G}\alpha = \mathcal{L} + G\alpha$ for all permissible $\tilde{F}$. This means that $\tilde{G}\alpha$ and $G\alpha$ are both elements of the coset $\mathcal{K}^*$, in fact $\tilde{G}\alpha = \mathcal{K}^* \cap \tilde{\mathcal{R}}$ and $G\alpha = \mathcal{K}^* \cap \mathcal{R}$.

In summary, in order to obtain the restricted least squares estimate of β we obtain the projection of the unrestricted least squares estimate b on $\mathcal{L}$ along $\mathcal{R}$, which is $Lb = (I - G\Psi)b$, and add to it the element $G\alpha \in \mathcal{R}$. Alternatively, we may introduce

<u>Definition 6.9.</u> Let $\mathcal{C} = \mathcal{E} + y$ be a coset in a linear space $\mathcal{J} = \mathcal{E} \oplus \bar{\mathcal{E}}$, where $y \in \bar{\mathcal{E}}$, and let $E : \mathcal{J} \rightarrow \mathcal{E} \mid \bar{\mathcal{E}}$. For any $x \in \mathcal{J}$ let P be the affine transformation defined by $Px = Ex + y$. The P is called the <u>affine projection</u> of $\mathcal{J}$ on $\mathcal{C} = \mathcal{E} + y$ along $\bar{\mathcal{E}}$, and Px is the affine projection of x on $\mathcal{C}$ along $\bar{\mathcal{E}}$.

Note that P is an idempotent transformation, for $y \in \bar{\mathcal{E}}$ hence $Ey = 0$ from Result 6.2, and so $PPx = E(Ex + y) + y = Ex + y = Px$. However P is not a linear transformation unless $y = 0$, since $P(u + v) = E(u + v) + y$ whereas $Pu + Pv = E(u + v) + 2y$.

Consequently $\hat{\beta}$ is the affine projection of b on $\mathcal{K}^* = \mathcal{L} + G\alpha$ along $\mathcal{R}$.

In considering two sets of restrictions $\Psi_0\beta = \alpha_0$ and $\Psi_1\beta = \alpha_1$, we may first obtain the estimate β^0 which takes account of the restrictions $\Psi_0\beta = \alpha_0$, and this is the affine projection P_0 of b on $\mathcal{L}_0 + G^0\alpha$ along $\mathcal{R}_0$, and is equal to $\beta^0 = L_0 b + G^0\alpha$. The affine projection P which takes account of both sets of restrictions is defined by $\hat{\beta} = Pb = Lb + G\alpha$. Since $\mathcal{L}_0$ is an invariant subspace under L (owing to the fact that $LL_0 = L_0 L = L$) we may obtain the same result by applying P to β^0 rather than b. Since $RG^0 = G\Psi G^0 = G^0$ from (5.16), it follows that $LG^0 = 0$, so $P\beta^0 = L(L_0 b + G^0\alpha) + G\alpha = Lb + G\alpha = \hat{\beta}$.

6.3. Perpendicular Projections in the Sample Space

We define

$$\begin{aligned} \mu &= X\beta \\ \hat{\mu} &= X\hat{\beta} \qquad (6.11) \\ m &= Xb \end{aligned}$$

and

$$\begin{aligned} u &= XG\Psi b \\ \nu &= XG\alpha \\ \hat{\delta} &= \hat{\mu} - \nu \qquad (6.12) \\ d &= m - \nu \end{aligned}$$

Then from (3.25) and (4.4),

$$\begin{aligned} \hat{\mu} &= m - u + \nu = m - h \\ \hat{\delta} &= m - h - \nu = m - u. \end{aligned}$$

From (4.5), (4.8), (4.9), (4.11), (6.12), and the definition $\hat{D} = D - H$ from (4.14),

$$\begin{aligned} Dz &= Dy - D\nu = m - \nu = d \\ Hz &= Hy - H\nu = u - \nu = h \qquad (6.14) \\ \hat{D}z &= Dz - Hz = d - h = \hat{\delta}. \end{aligned}$$

Now from (4.14) and (5.12),

$$I = D + E = \hat{D} + H + E = \hat{D} + H_0 + H\# + E \qquad (6.15)$$

so that, from (4.10) and (5.21),

$$(6.16)\qquad z = d + e = \hat{\delta} + h + e = \hat{\delta} + h_0 + h\# + e$$

(compare Bartlett [4, pp. 328-9]). The matrices A of (6.15) all satisfy the properties of Theorem 1.8, i.e., each A is idempotent and $A\Omega$ is symmetric; furthermore $A'\Omega^{-1}A = \Omega^{-1}A$. Thus we have the decomposition into quadratic forms

$$(6.17)\qquad z'\Omega^{-1}z = d'\Omega^{-1}d + e'\Omega^{-1}e = \hat{\delta}'\Omega^{-1}\hat{\delta} + h'\Omega^{-1}h + e'\Omega^{-1}e$$
$$= \hat{\delta}'\Omega^{-1}\delta + h^{0\prime}\Omega^{-1}h^0 + h\#'\Omega^{-1}h\# + e'\Omega^{-1}e$$

of ranks

$$(6.18)\qquad n = k + (n - k) = (k - r) + r + (n - k) = (k - r) + r_0 + (r - r_0) + (n-k)$$

From Theorem 1.9 (Cochran's Theorem) these are independently distributed by chi-square laws. If $\Omega = I$ then the matrices in (6.15) are all symmetric, hence the corresponding projections are perpendicular. (6.17) is then a division of z'z into sums of squares, and (6.17) is simply a statement of the Pythagorean theorem, since $\sqrt{z'z}$ is the length of the vector z.

In Figure 2 an illustration is given for the case

$$(6.18)\qquad X = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \qquad y = \begin{bmatrix} 6 \\ 1 \end{bmatrix} \qquad \begin{matrix} \Psi = 1 \\ \alpha = 2 \end{matrix} \qquad \Omega = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$$

From (6.18) it follows that

$$(6.19)\qquad D = H = \begin{bmatrix} \frac{1}{2} & \frac{1}{2} \\ \frac{1}{2} & \frac{1}{2} \end{bmatrix} \qquad E = J = \begin{bmatrix} \frac{1}{2} & -\frac{1}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{bmatrix}$$

For all $\mu \in \mathcal{D}$, the distance between y and μ is $\sqrt{\varepsilon'\varepsilon}$ where $\varepsilon = y - \mu$. This distance is minimized when the projection D: $y \to \mathcal{D} \mid \mathcal{E}$ is perpendicular, which is the case when $\mu = m$ and $\varepsilon = e$, i.e., when the estimator for μ is the least squares estimator. Now Bartlett has shown [4, p.331] that a vector y is distributed as $N(\mu, \sigma^2 I)$ if and only if the probability density of y depends only on the length $\sqrt{y'y}$. Thus the locus of constant density is given by the circumference of a hypersphere around any $\mu \in \mathcal{D}$ (in Figure 2, two circles of equal radius are shown with centers m and d). If we take the family of hyper-

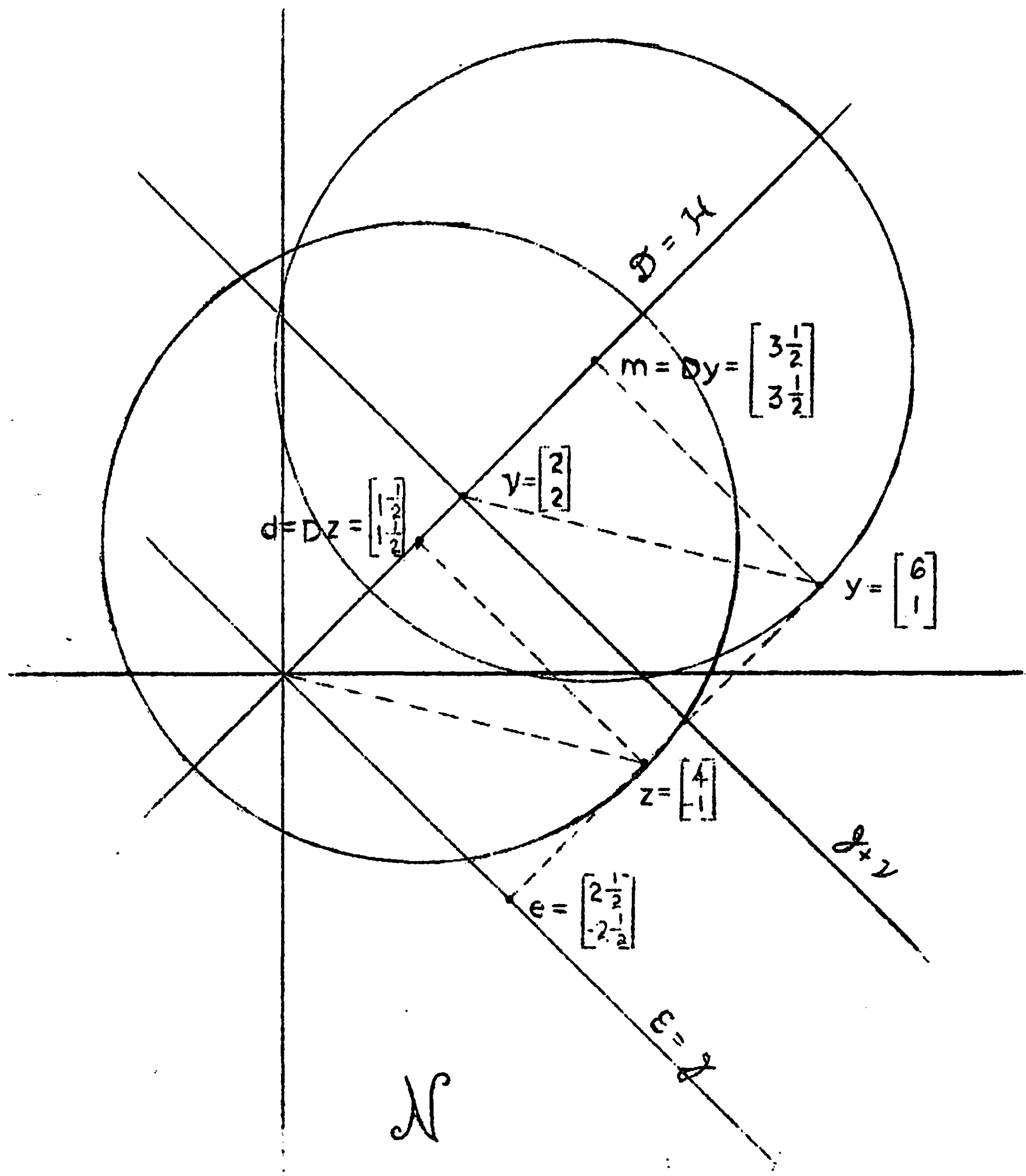

$\mathcal{D} = \mathcal{H}$
$m = \mathcal{D}y = \begin{bmatrix} 3\frac{1}{2} \\ 3\frac{1}{2} \end{bmatrix}$
$\gamma = \begin{bmatrix} 2 \\ 2 \end{bmatrix}$
$d = \mathcal{D}z = \begin{bmatrix} 1\frac{1}{2} \\ 1\frac{1}{2} \end{bmatrix}$
$y = \begin{bmatrix} 6 \\ 1 \end{bmatrix}$
$z = \begin{bmatrix} 4 \\ -1 \end{bmatrix}$
$e = \begin{bmatrix} 2\frac{1}{2} \\ -2\frac{1}{2} \end{bmatrix}$
$\mathcal{J} + \gamma$
$\mathcal{E} = \mathcal{L}$
$\mathcal{N}$

Figure 2

spheres with centers $\mu \in \mathscr{D}$ and circumferences all passing through y, that hypersphere with minimum radius has m as its center; thus if $y \sim N(\mu, \sigma^2 I)$, m is the maximum likelihood estimate of μ. If $y \sim N(\mu, \sigma^2 \Omega)$, the hyperspheres in the above analysis will be replaced by ellipsoids. Corresponding to a given Ω there is a subspace $\mathcal{E}$; the quotient set $\mathcal{N}/\mathcal{E}$, defined as the set of all cosets $\delta + \mathcal{E}$ where $\delta \in \mathscr{D}$, is a sufficient partition of $\mathcal{N}$.

From (6.13) and (6.14) we have

$$\hat{\mu} = m - u + \nu = Dy - Hy + \nu = \hat{D}y + \nu \tag{6.20}$$

since $\hat{D} = D - H$. Since $D - H = (I - H)\ D = JD$, this becomes

$$\hat{\mu} = JDy + \nu = Jm + \nu \tag{6.21}$$

which corresponds to the result

$$\hat{\beta} = LBy + \rho = Lb + \rho \tag{6.22}$$

from (3.25), where $\rho = G\alpha$ and B is defined by (5.14).

REFERENCES

[1] A. C. Aitken, Determinants and Matrices, Edinburgh, Oliver and Boyd, 1948.

[2] A. C. Aitken, "On least squares and linear combinations of observations," Proceedings of the Royal Society of Edinburgh, 55 (1934), 42-48.

[3] A. C. Aitken, "On the statistical independence of quadratic forms in normal variates," Biometrika, 37 (1950), 93-96.

[4] M. S. Bartlett, "The vector representation of a sample," Proceedings of the Cambridge Philosophical Society, 30 (1934), 327-340.

[5] G. E. P. Box and J. S. Hunter, "Multi-factor experimental designs for exploring response surfaces," Annals of Mathematical Statistics, 28 (1957), 195-241.

[6] G. E. P. Box and K. B. Wilson, "On the experimental attainment of optimum conditions," Journal of the Royal Statistical Society (Series B), 13 (1951), 1-45.

[7] Osmer Carpenter, "Note on the extension of Craig's theorem to non-central variates," Annals of Mathematical Statistics, 21 (1950), 455-457.

[8] W. G. Cochran, "The distribution of quadratic forms in a normal system, with application to the analysis of covariance," Proceedings of the Cambridge Philosophical Society, 30, (1934), 178-191.

[9] Allen T. Craig, "Note on the independence of certain quadratic forms," Annals of Mathematical Statistics, 14 (1943), 195-197.

[10] Allen T. Craig, "Bilinear forms in normally correlated variables," Annals of Mathematical Statistics, 18 (1947), 565-573.

[11] Harald Cramer, Mathematical Methods of Statistics, Princeton, Princeton University Press, 1946.

[12] F. N. David and J. Neyman, "Extension of the Markoff theorem on least squares," Statistical Research Memoirs, 2 (1938), 105-116.

[13] Paul S. Dwyer, "Generalizations of a Gaussian theorem," Annals of Mathematical Statistics, 29 (1958), 106-117.

[14] Franklin A. Graybill and George Marsaglia, "Idempotent matrices and quadratic forms in the general linear hypothesis," Annals of Mathematical Statistics, 28 (1957), 678-686.

[15] Paul R. Halmos, Finite-Dimensional Vector Spaces, Second Edition, New York, D. Van Nostrand Company, Inc., 1958.

[16] Robert V. Hogg and Allen T. Craig, "On the decomposition of certain χ^2 variables," Annals of Mathematical Statistics, 29 (1958), 608-610.

[17] Harold Hotelling, "Note on a matric theorem of A. T. Craig," Annals of Mathematical Statistics, 15 (1944), 427-429.

[18] Nathan Jacobson, Lectures in Abstract Algebra, New York, D. Van Nostrand Company, Inc., Vol. I, 1958, Vol. II, 1953.

[19] T. C. Koopmans, H. Rubin, and R. B. Leipnik, "Measuring the equation systems of dynamic economics," Chapter II in Statistical Inference in Dynamic Economic Models, edited by Tjalling C. Koopmans, New York, John Wiley and Sons, Inc., 1950.

[20] William G. Madow, "The distribution of quadratic forms in non-central normal random variables," Annals of Mathematical Statistics, 11 (1940), 100-103.

[21] Bertil Matérn, "Independence of non-negative quadratic forms in normally correlated variables," Annals of Mathematical Statistics, 20 (1949), 119-120.

[22] R. D. Narain, "On the completely unbiased character of tests of independence in multivariate normal systems," Annals of Mathematical Statistics, 21 (1950), 293-298.

[23] R. L. Plackett, "A historical note on the method of least squares," Biometrika, 36 (1949), 458-460.

[24] C. Radhakrishna Rao, "Generalization of Markoff's theorem and tests of linear hypotheses," Sankhyā, 7 (1945), 9-16.

[25] C. Radhakrishna Rao, "Markoff's theorem with linear restrictions on parameters," Sankhyā, 7 (1945), 16-19.

[26] K. V. Ramachandran, "On the Tukey test for the equality of means and the Hartley test for the equality of variances," Annals of Mathematical Statistics, 27 (1956), 825-831.

[27] Richard Stone, assisted by D. A. Rowe and by W. J. Corlett, Renee Hurstfield and Muriel Potter, The Measurement of Consumers' Expenditure and Behaviour in the United Kingdom, 1920-1938, Vol. I, Cambridge University Press, 1954.

[28] P. C. Tang, "The power function of the analysis of variance tests with tables and illustrations of their use," Statistical Research Memoirs, 2 (1938), 126-157.

[29] H. Theil, Linear Aggregation of Economic Relations, Amsterdam, North-Holland Publishing Company, 1954.

[30] H. Theil, "Specification errors and the estimation of economic relationships," Revue de l'Institut International de Statistique, 25 (1957), 41-51.

[31] S. S. Wilks, Mathematical Statistics, Princeton University Press, 1947.

Department of Statistics
University of Minnesota

BASIC DISTRIBUTION LIST FOR UNCLASSIFIED TECHNICAL REPORTS

for contract Nonr 2582(00), Task NR 042-200

ADDRESS

Head, Statistics Branch
Office of Naval Research
Washington 25, D. C. 2

Commanding Officer
Office of Naval Research Branch Office
Navy Nr. 100
Fleet Post Office
New York, New York 2

ASTIA Document Service Center
Arlington Hall Station
Arlington 12, Virginia 10

Office of Technical Services
Department of Commerce
Washington 25, D. C. 1

Technical Information Officer
Naval Research Laboratory
Washington 25, D. C. 6

Professor T. W. Anderson
Department of Mathematical Statistics
Columbia University
New York 27, New York 1

Professor Z. W. Birnbaum
Laboratory of Statistical Research
Department of Mathematics
University of Washington
Seattle 5, Washington 1

Professor A. H. Bowker
Applied Mathematics and Statistics Lab.
Stanford University
Stanford, California 1

Professor Ralph A. Bradley
Department of Statistics and
Statistical Laboratory
Virginia Polytechnic Institute
Blacksburg, Virginia 1

Professor Herman Chernoff
Applied Mathematics and Statistics Lab.
Stanford University
Stanford, California 1

ADDRESS

Professor W.G. Cochran
Department of Statistics
Harvard University
Cambridge, Massachusetts 1

Professor Benjamin Epstein
Applied Mathematics and Statistics Lab.
Stanford University
Stanford, California 1

Professor W. Hirsch
Institute of Mathematical Sciences
New York University
New York 3, New York 1

Dr. Paul G. Hoel
Department of Mathematics
University of California
Los Angeles 24, California 1

Professor Harold Hotelling
Associate Director
Institute of Statistics
University of North Carolina
Chapel Hill, North Carolina 1

Professor L. Hurwicz
School of Business Administration
University of Minnesota
Minneapolis 14, Minnesota 1

Professor Leo Katz
Department of Statistics
Michigan State University
East Lansing, Michigan 1

Professor Oscar Kempthorne
Statistics Laboratory
Iowa State College
Ames, Iowa 1

Dr. Carl F. Kossack
Statistical Laboratory
Engineering Administration Building
Purdue University
Lafayette, Indiana

ADDRESS		ADDRESS	
Professor Gerald J. Lieberman Applied Mathematics and Statistics Laboratory Stanford University Stanford, California	1	Professor Evan J. Williams Institute of Statistics State College Section North Carolina State College Raleigh, North Carolina	1
Professor William G. Madow Department of Statistics Stanford University Stanford, California	1	Professor J. Wolfowitz Department of Mathematics Cornell University Ithaca, New York	1
Professor J. Neyman Department of Statistics University of California Berkeley 4, California	1		
Professor Herbert Robbins Mathematical Statistics Department Columbia University New York 27, New York	1		
Professor Murray Rosenblatt Department of Mathematics Indiana University Bloomington, Indiana	1		
Professor L. J. Savage Statistical Research Laboratory Chicago University Chicago 37, Illinois	1		
Dr. Herbert Solomon Teachers College Department of Statistics Columbia University New York 27, New York	1		
Professor Frank Spitzer Department of Mathematics University of Minnesota Minneapolis 14, Minnesota	1		
Professor S. S. Wilks Department of Mathematics Princeton University Princeton, New Jersey	1		